(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
24 January 2002 (24.01.2002)

PCT

(10) International Publication Number
WO 02/06339 A2

(51) International Patent Classification[7]: C07K 14/47

(21) International Application Number: PCT/US01/21249

(22) International Filing Date: 3 July 2001 (03.07.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/215,854	3 July 2000 (03.07.2000)	US
60/215,856	3 July 2000 (03.07.2000)	US
60/215,902	3 July 2000 (03.07.2000)	US
60/216,585	7 July 2000 (07.07.2000)	US
60/216,586	7 July 2000 (07.07.2000)	US
60/216,722	7 July 2000 (07.07.2000)	US
60/218,622	17 July 2000 (17.07.2000)	US
60/218,992	17 July 2000 (17.07.2000)	US
60/221,285	27 July 2000 (27.07.2000)	US
60/268,734	14 February 2001 (14.02.2001)	US
60/274,260	8 March 2001 (08.03.2001)	US
60/279,856	29 March 2001 (29.03.2001)	US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier applications:

US	60/215,854 (CIP)
Filed on	3 July 2000 (03.07.2000)
US	60/215,856 (CIP)
Filed on	3 July 2000 (03.07.2000)
US	60/215,902 (CIP)
Filed on	3 July 2000 (03.07.2000)
US	60/216,585 (CIP)
Filed on	7 July 2000 (07.07.2000)
US	60/216,586 (CIP)
Filed on	7 July 2000 (07.07.2000)
US	60/216,722 (CIP)
Filed on	7 July 2000 (07.07.2000)
US	60/218,622 (CIP)
Filed on	17 July 2000 (17.07.2000)
US	60/218,992 (CIP)
Filed on	17 July 2000 (17.07.2000)
US	60/221,285 (CIP)
Filed on	27 July 2000 (27.07.2000)
US	60/268,734 (CIP)
Filed on	14 February 2001 (14.02.2001)
US	60/274,260 (CIP)
Filed on	8 March 2001 (08.03.2001)
US	60/279,856 (CIP)
Filed on	29 March 2001 (29.03.2001)

(71) Applicant *(for all designated States except US)*: **CURAGEN CORPORATION** [US/US]; 555 Long Wharf Drive, 11th Floor, New Haven, CT 06511 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **SPADERNA, Steven, K.** [US/US]; 261 Deerfield Drive, Berlin, CT 06037 (US). **TCHERNEV, Velizar** [BG/US]; 45 Jefferson Road #3-12, Branford, CT 06405 (US). **LIU, Xiaohong** [CN/US]; 96 Montoya Circle, Branford, CT 06405 (US). **SHENOY, Suresh** [IN/US]; 15 Millwood Drive, Branford, CT 06405 (US). **SPYTEK, Kimberly** [US/US]; 28 Court Street #1, New Haven, CT 06511 (US). **ZERHUSEN, Bryan** [US/US]; 45 Jefferson Road, Apartment #2-10, Branford, CT 06405 (US). **PATTURAJAN, Meera** [IN/US]; 45 Harrison Avenue, Apartment 1C, Branford, CT 06405 (US). **TAUPIER, Raymond, J.** [US/US]; 47 Holmes Street, East Haven, CT 06512 (US). **RASTELLI, Luca** [IT/US]; 52 Pepperbush Lane, Guilford, CT 06437 (US). **GROSSE, William, M.** [US/US]; 15 Rice Terrace Road, Branford, CT 06405 (US). **SZEKERES, Edward, S.** [US/US]; 130 Montoya Drive, Branford, CT 06504 (US). **ALSOBROOK, John, II** [US/US]; 60 Lake Drive, Madison, CT 06443 (US). **LEPLEY, Denise, M.** [US/US]; 322 East Main Strait, Branford, CT 06405 (US). **SHEN, Lei** [US/US]; 322 East Main Street, Branford, CT 06405 (US). **BURGESS, Catherine, E.** [US/US]; 90 Carriage Hill Drive, Wethersfield, CT 06109 (US). **SHIMKETS, Richard, A.** [US/US]; 191 Leete Street, New Haven, CT 06516 (US). **PADIGARU, Muralidhara** [IN/US]; 98 Montoya Circle, Branford, CT 06405 (US).

(74) Agent: **ELRIFI, Ivor, R.**; Mintz, Levin, Cohn, Ferris, Glovsky and Popeo, P.C., One Financial Center, Boston, MA 02111 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX,

[Continued on next page]

(54) Title: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME

(57) Abstract: Disclosed herein are nucleic acid sequences that encode novel polypeptides. Also disclosed are polypeptides encoded by these nucleic acid sequences, and antibodies, which immunospecifically-bind to the polpeptide, as well as derivatives, variants, mutants, or fragments of the aforementioned polypeptide, polynucleotide, or antibody. The invention further discloses therapeutic, diagnostic and research methods for diagnosis, treatment, and prevention of disorders involving any one of these novel human nucleic acids and proteins.

WO 02/06339 A2

MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) **Designated States** *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:

— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME

BACKGROUND OF THE INVENTION

The invention generally relates to nucleic acids and polypeptides. More particularly, the invention relates to nucleic acids encoding novel molecule (MOL) polypeptides, as well as vectors, host cells, antibodies, and recombinant methods for producing these nucleic acids and polypeptides.

SUMMARY OF THE INVENTION

The invention is based in part upon the discovery of nucleic acid sequences encoding novel polypeptides. The novel nucleic acids and polypeptides are referred to herein as MOLX, or MOL1, MOL2, MOL3, MOL4, MOL5, MOL6, MOL7, and MOL8 nucleic acids and polypeptides. These nucleic acids and polypeptides, as well as derivatives, homologs, analogs and fragments thereof, will hereinafter be collectively designated as "MOLX" nucleic acid or polypeptide sequences.

In one aspect, the invention provides an isolated MOLX nucleic acid molecule encoding a MOLX polypeptide that includes a nucleic acid sequence that has identity to the nucleic acids disclosed in SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, and 23. In some embodiments, the MOLX nucleic acid molecule will hybridize under stringent conditions to a nucleic acid sequence complementary to a nucleic acid molecule that includes a protein-coding sequence of a MOLX nucleic acid sequence. The invention also includes an isolated nucleic acid that encodes a MOLX polypeptide, or a fragment, homolog, analog or derivative thereof. For example, the nucleic acid can encode a polypeptide at least 80% identical to a polypeptide comprising the amino acid sequences of SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, and 24. The nucleic acid can be, for example, a genomic DNA fragment or a cDNA molecule that includes the nucleic acid sequence of any of SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, and 23.

Also included in the invention is an oligonucleotide, *e.g.*, an oligonucleotide which includes at least 6 contiguous nucleotides of a MOLX nucleic acid (*e.g.*, SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, and 23) or a complement of said oligonucleotide.

Also included in the invention are substantially purified MOLX polypeptides (SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, and 24). In certain embodiments, the MOLX

polypeptides include an amino acid sequence that is substantially identical to the amino acid sequence of a human MOLX polypeptide.

The invention also features antibodies that immunoselectively bind to MOLX polypeptides, or fragments, homologs, analogs or derivatives thereof.

In another aspect, the invention includes pharmaceutical compositions that include therapeutically- or prophylactically-effective amounts of a therapeutic and a pharmaceutically-acceptable carrier. The therapeutic can be, *e.g.*, a MOLX nucleic acid, a MOLX polypeptide, or an antibody specific for a MOLX polypeptide. In a further aspect, the invention includes, in one or more containers, a therapeutically- or prophylactically-effective amount of this pharmaceutical composition.

In a further aspect, the invention includes a method of producing a polypeptide by culturing a cell that includes a MOLX nucleic acid, under conditions allowing for expression of the MOLX polypeptide encoded by the DNA. If desired, the MOLX polypeptide can then be recovered.

In another aspect, the invention includes a method of detecting the presence of a MOLX polypeptide in a sample. In the method, a sample is contacted with a compound that selectively binds to the polypeptide under conditions allowing for formation of a complex between the polypeptide and the compound. The complex is detected, if present, thereby identifying the MOLX polypeptide within the sample.

The invention also includes methods to identify specific cell or tissue types based on their expression of a MOLX.

Also included in the invention is a method of detecting the presence of a MOLX nucleic acid molecule in a sample by contacting the sample with a MOLX nucleic acid probe or primer, and detecting whether the nucleic acid probe or primer bound to a MOLX nucleic acid molecule in the sample.

In a further aspect, the invention provides a method for modulating the activity of a MOLX polypeptide by contacting a cell sample that includes the MOLX polypeptide with a compound that binds to the MOLX polypeptide in an amount sufficient to modulate the activity of said polypeptide. The compound can be, *e.g.*, a small molecule, such as a nucleic acid, peptide, polypeptide, peptidomimetic, carbohydrate, lipid or other organic (carbon-containing) or inorganic molecule, as further described herein.

Also within the scope of the invention is the use of a therapeutic in the manufacture of a medicament for treating or preventing disorders or syndromes including, *e.g.*, endometriosis, fertility disorders, hypercoagulation, idiopathic thrombocytopenic purpura,

immunodeficiencies, systemic lupus erythematosus, asthma, emphysema, scleroderma, allergy, ARDS, Von Hippel-Lindau (VHL) syndrome, Alzheimer's disease, stroke, tuberous sclerosis, hypercalceimia, Parkinson's disease, Huntington's disease, cerebral palsy, epilepsy, multiple sclerosis, ulcers, ataxia-telangiectasia, leukodystrophies, behavioral disorders, addiction, anxiety, pain, neuroprotection, osteoporosis, arthritis, ankylosing spondylitis, scoliosis, diabetes, autoimmune disease, myasthenia gravis, muscular dystrophy, renal artery stenosis, interstitial nephritis, glomerulonephritis, polycystic kidney disease, systemic lupus erythematosus, renal tubular acidosis, IgA nephropathy, Lesch-Nyhan syndrome, developmental disorders, growth disorders, and/or wounds, cardiomyopathy, atherosclerosis, hypertension, congenital heart defects, aortic stenosis, atrial septal defect (ASD), atrioventricular (A-V) canal defect, ductus arteriosus , pulmonary stenosis , subaortic stenosis, ventricular septal defect (VSD), valve diseases, obesity, transplantation, hemophilia, hypercoagulation, idiopathic thrombocytopenic purpura, immunodeficiencies, graft versus host disease (GVHD), lymphaedema, adrenoleukodystrophy, congenital adrenal hyperplasia, neuronal developmental, organizational, mediated and interactive disorders and disease; endocrine dysfunctions, growth and reproductive disorders, injury repair, cancer including but not limited to lung or breast cancer, endocrine disorders, inflammatory disorders, gastro-intestinal disorders and disorders of the respiratory system, Rheumatoid arthritis (RA), CNS disorders, Down syndrome, Schizophrenia, nutritional deficiencies, primary open-angle glaucoma (POAG), and bone disorders, hematopoietic disorders, or other disorders. The therapeutic can be, *e.g.*, a MOLX nucleic acid, a MOLX polypeptide, or a MOLX-specific antibody, or biologically-active derivatives or fragments thereof.

For example, the compositions of the present invention will have efficacy for treatment of patients suffering from: Cancer including endometriosis, fertility disorders, hypercoagulation, idiopathic thrombocytopenic purpura, immunodeficiencies, systemic lupus erythematosus, asthma, emphysema, scleroderma, allergy, ARDS, Von Hippel-Lindau (VHL) syndrome, Alzheimer's disease, stroke, tuberous sclerosis, hypercalceimia, Parkinson's disease, Huntington's disease, cerebral palsy, epilepsy, multiple sclerosis, ulcers, ataxia-telangiectasia, leukodystrophies, behavioral disorders, addiction, anxiety, pain, neuroprotection, osteoporosis, arthritis, ankylosing spondylitis, scoliosis, diabetes, autoimmune disease, myasthenia gravis, muscular dystrophy, renal artery stenosis, interstitial nephritis, glomerulonephritis, polycystic kidney disease, systemic lupus erythematosus, renal tubular acidosis, IgA nephropathy, Lesch-Nyhan syndrome, developmental disorders, growth disorders, and/or wounds, cardiomyopathy, atherosclerosis, hypertension, congenital heart defects, aortic stenosis, atrial

septal defect (ASD), atrioventricular (A-V) canal defect, ductus arteriosus , pulmonary stenosis , subaortic stenosis, ventricular septal defect (VSD), valve diseases, obesity, transplantation, hemophilia, hypercoagulation, idiopathic thrombocytopenic purpura, immunodeficiencies, graft versus host disease (GVHD), lymphaedema, adrenoleukodystrophy, congenital adrenal hyperplasia, neuronal developmental, organizational, mediated and interactive disorders and disease; endocrine dysfunctions, growth and reproductive disorders, injury repair, cancer including but not limited to lung or breast cancer, endocrine disorders, inflammatory disorders, gastro-intestinal disorders and disorders of the respiratory system, Rheumatoid arthritis (RA), CNS disorders, Down syndrome, Schizophrenia, nutritional deficiencies, primary open-angle glaucoma (POAG), and bone disorders, hematopoietic disorders and/or other pathologies and disorders of the like.

The polypeptides can be used as immunogens to produce antibodies specific for the invention and as vaccines. They can also be used to screen for potential agonist and antagonist compounds. For example, a cDNA encoding MOLX may be useful in gene therapy, and MOLX may be useful when administered to a subject in need thereof. By way of nonlimiting example, the compositions of the present invention will have efficacy for treatment of patients suffering from endometriosis, fertility disorders, hypercoagulation, idiopathic thrombocytopenic purpura, immunodeficiencies, systemic lupus erythematosus, asthma, emphysema, scleroderma, allergy, ARDS, Von Hippel-Lindau (VHL) syndrome, Alzheimer's disease, stroke, tuberous sclerosis, hypercalceimia, Parkinson's disease, Huntington's disease, cerebral palsy, epilepsy, multiple sclerosis, ulcers, ataxia-telangiectasia, leukodystrophies, behavioral disorders, addiction, anxiety, pain, neuroprotection, osteoporosis, arthritis, ankylosing spondylitis, scoliosis, diabetes, autoimmune disease, myasthenia gravis, muscular dystrophy, renal artery stenosis, interstitial nephritis, glomerulonephritis, polycystic kidney disease, systemic lupus erythematosus, renal tubular acidosis, IgA nephropathy, Lesch-Nyhan syndrome, developmental disorders, growth disorders, and/or wounds, cardiomyopathy, atherosclerosis, hypertension, congenital heart defects, aortic stenosis, atrial septal defect (ASD), atrioventricular (A-V) canal defect, ductus arteriosus , pulmonary stenosis , subaortic stenosis, ventricular septal defect (VSD), valve diseases, obesity, transplantation, hemophilia, hypercoagulation, idiopathic thrombocytopenic purpura, immunodeficiencies, graft versus host disease (GVHD), lymphaedema, adrenoleukodystrophy, congenital adrenal hyperplasia, neuronal developmental, organizational, mediated and interactive disorders and disease; endocrine dysfunctions, growth and reproductive disorders, injury repair, cancer including but not limited to lung or breast cancer, endocrine disorders, inflammatory disorders, gastro-

intestinal disorders and disorders of the respiratory system, Rheumatoid arthritis (RA), CNS disorders, Down syndrome, Schizophrenia, nutritional deficiencies, primary open-angle glaucoma (POAG), and bone disorders, hematopoietic disorders and/or other pathologies and disorders.

The invention further includes a method for screening for a modulator of disorders or syndromes including, *e.g.*, endometriosis, fertility disorders, hypercoagulation, idiopathic thrombocytopenic purpura, immunodeficiencies, systemic lupus erythematosus, asthma, emphysema, scleroderma, allergy, ARDS, Von Hippel-Lindau (VHL) syndrome, Alzheimer's disease, stroke, tuberous sclerosis, hypercalceimia, Parkinson's disease, Huntington's disease, cerebral palsy, epilepsy, multiple sclerosis, ulcers, ataxia-telangiectasia, leukodystrophies, behavioral disorders, addiction, anxiety, pain, neuroprotection, osteoporosis, arthritis, ankylosing spondylitis, scoliosis, diabetes, autoimmune disease, myasthenia gravis, muscular dystrophy, renal artery stenosis, interstitial nephritis, glomerulonephritis, polycystic kidney disease, systemic lupus erythematosus, renal tubular acidosis, IgA nephropathy, Lesch-Nyhan syndrome, developmental disorders, growth disorders, and/or wounds, cardiomyopathy, atherosclerosis, hypertension, congenital heart defects, aortic stenosis, atrial septal defect (ASD), atrioventricular (A-V) canal defect, ductus arteriosus , pulmonary stenosis , subaortic stenosis, ventricular septal defect (VSD), valve diseases, obesity, transplantation, hemophilia, hypercoagulation, idiopathic thrombocytopenic purpura, immunodeficiencies, graft versus host disease (GVHD), lymphaedema, adrenoleukodystrophy, congenital adrenal hyperplasia, neuronal developmental, organizational, mediated and interactive disorders and disease; endocrine dysfunctions, growth and reproductive disorders, injury repair, cancer including but not limited to lung or breast cancer, endocrine disorders, inflammatory disorders, gastro-intestinal disorders and disorders of the respiratory system, Rheumatoid arthritis (RA), CNS disorders, Down syndrome, Schizophrenia, nutritional deficiencies, primary open-angle glaucoma (POAG), and bone disorders, hematopoietic disorders or other disorders related to cell signal processing and metabolic pathway modulation. The method includes contacting a test compound with a MOLX polypeptide and determining if the test compound binds to said MOLX polypeptide. Binding of the test compound to the MOLX polypeptide indicates the test compound is a modulator of activity, or of latency or predisposition to the aforementioned disorders or syndromes.

Also within the scope of the invention is a method for screening for a modulator of activity, or of latency or predisposition to an disorders or syndromes including, *e.g.*, endometriosis, fertility disorders, hypercoagulation, idiopathic thrombocytopenic purpura,

immunodeficiencies, systemic lupus erythematosus, asthma, emphysema, scleroderma, allergy, ARDS, Von Hippel-Lindau (VHL) syndrome, Alzheimer's disease, stroke, tuberous sclerosis, hypercalceimia, Parkinson's disease, Huntington's disease, cerebral palsy, epilepsy, multiple sclerosis, ulcers, ataxia-telangiectasia, leukodystrophies, behavioral disorders, addiction, anxiety, pain, neuroprotection, osteoporosis, arthritis, ankylosing spondylitis, scoliosis, diabetes, autoimmune disease, myasthenia gravis, muscular dystrophy, renal artery stenosis, interstitial nephritis, glomerulonephritis, polycystic kidney disease, systemic lupus erythematosus, renal tubular acidosis, IgA nephropathy, Lesch-Nyhan syndrome, developmental disorders, growth disorders, and/or wounds, cardiomyopathy, atherosclerosis, hypertension, congenital heart defects, aortic stenosis, atrial septal defect (ASD), atrioventricular (A-V) canal defect, ductus arteriosus , pulmonary stenosis , subaortic stenosis, ventricular septal defect (VSD), valve diseases, obesity, transplantation, hemophilia, hypercoagulation, idiopathic thrombocytopenic purpura, immunodeficiencies, graft versus host disease (GVHD), lymphaedema, adrenoleukodystrophy, congenital adrenal hyperplasia, neuronal developmental, organizational, mediated and interactive disorders and disease; endocrine dysfunctions, growth and reproductive disorders, injury repair, cancer including but not limited to lung or breast cancer, endocrine disorders, inflammatory disorders, gastro-intestinal disorders and disorders of the respiratory system, Rheumatoid arthritis (RA), CNS disorders, Down syndrome, Schizophrenia, nutritional deficiencies, primary open-angle glaucoma (POAG), and bone disorders, hematopoietic disorders or other disorders related to cell signal processing and metabolic pathway modulation by administering a test compound to a test animal at increased risk for the aforementioned disorders or syndromes. The test animal expresses a recombinant polypeptide encoded by a MOLX nucleic acid. Expression or activity of MOLX polypeptide is then measured in the test animal, as is expression or activity of the protein in a control animal which recombinantly-expresses MOLX polypeptide and is not at increased risk for the disorder or syndrome. Next, the expression of MOLX polypeptide in both the test animal and the control animal is compared. A change in the activity of MOLX polypeptide in the test animal relative to the control animal indicates the test compound is a modulator of latency of the disorder or syndrome.

In yet another aspect, the invention includes a method for determining the presence of or predisposition to a disease associated with altered levels of a MOLX polypeptide, a MOLX nucleic acid, or both, in a subject (*e.g.*, a human subject). The method includes measuring the amount of the MOLX polypeptide in a test sample from the subject and comparing the amount of the polypeptide in the test sample to the amount of the MOLX polypeptide present in a

control sample. An alteration in the level of the MOLX polypeptide in the test sample as compared to the control sample indicates the presence of or predisposition to a disease in the subject. Preferably, the predisposition includes, *e.g.*, endometriosis, fertility disorders, hypercoagulation, idiopathic thrombocytopenic purpura, immunodeficiencies, systemic lupus erythematosus, asthma, emphysema, scleroderma, allergy, ARDS, Von Hippel-Lindau (VHL) syndrome, Alzheimer's disease, stroke, tuberous sclerosis, hypercalceimia, Parkinson's disease, Huntington's disease, cerebral palsy, epilepsy, multiple sclerosis, ulcers, ataxia-telangiectasia, leukodystrophies, behavioral disorders, addiction, anxiety, pain, neuroprotection, osteoporosis, arthritis, ankylosing spondylitis, scoliosis, diabetes, autoimmune disease, myasthenia gravis, muscular dystrophy, renal artery stenosis, interstitial nephritis, glomerulonephritis, polycystic kidney disease, systemic lupus erythematosus, renal tubular acidosis, IgA nephropathy, Lesch-Nyhan syndrome, developmental disorders, growth disorders, and/or wounds, cardiomyopathy, atherosclerosis, hypertension, congenital heart defects, aortic stenosis, atrial septal defect (ASD), atrioventricular (A-V) canal defect, ductus arteriosus , pulmonary stenosis , subaortic stenosis, ventricular septal defect (VSD), valve diseases, obesity, transplantation, hemophilia, hypercoagulation, idiopathic thrombocytopenic purpura, immunodeficiencies, graft versus host disease (GVHD), lymphaedema, adrenoleukodystrophy, congenital adrenal hyperplasia, neuronal developmental, organizational, mediated and interactive disorders and disease; endocrine dysfunctions, growth and reproductive disorders, injury repair, cancer including but not limited to lung or breast cancer, endocrine disorders, inflammatory disorders, gastro-intestinal disorders and disorders of the respiratory system, Rheumatoid arthritis (RA), CNS disorders, Down syndrome, Schizophrenia, nutritional deficiencies, primary open-angle glaucoma (POAG); and bone disorders, hematopoietic disorders. Also, the expression levels of the new polypeptides of the invention can be used in a method to screen for various cancers as well as to determine the stage of cancers.

In a further aspect, the invention includes a method of treating or preventing a pathological condition associated with a disorder in a mammal by administering to the subject a MOLX polypeptide, a MOLX nucleic acid, or a MOLX-specific antibody to a subject (*e.g.*, a human subject), in an amount sufficient to alleviate or prevent the pathological condition. In preferred embodiments, the disorder, includes, *e.g.*, endometriosis, fertility disorders, hypercoagulation, idiopathic thrombocytopenic purpura, immunodeficiencies, systemic lupus erythematosus, asthma, emphysema, scleroderma, allergy, ARDS, Von Hippel-Lindau (VHL) syndrome, Alzheimer's disease, stroke, tuberous sclerosis, hypercalceimia, Parkinson's disease, Huntington's disease, cerebral palsy, epilepsy, multiple sclerosis, ulcers, ataxia-telangiectasia,

leukodystrophies, behavioral disorders, addiction, anxiety, pain, neuroprotection, osteoporosis, arthritis, ankylosing spondylitis, scoliosis, diabetes, autoimmune disease, myasthenia gravis, muscular dystrophy, renal artery stenosis, interstitial nephritis, glomerulonephritis, polycystic kidney disease, systemic lupus erythematosus, renal tubular acidosis, IgA nephropathy, Lesch-Nyhan syndrome, developmental disorders, growth disorders, and/or wounds, cardiomyopathy, atherosclerosis, hypertension, congenital heart defects, aortic stenosis, atrial septal defect (ASD), atrioventricular (A-V) canal defect, ductus arteriosus , pulmonary stenosis , subaortic stenosis, ventricular septal defect (VSD), valve diseases, obesity, transplantation, hemophilia, hypercoagulation, idiopathic thrombocytopenic purpura, immunodeficiencies, graft versus host disease (GVHD), lymphaedema, adrenoleukodystrophy, congenital adrenal hyperplasia, neuronal developmental, organizational, mediated and interactive disorders and disease; endocrine dysfunctions, growth and reproductive disorders, injury repair, cancer including but not limited to lung or breast cancer, endocrine disorders, inflammatory disorders, gastro-intestinal disorders and disorders of the respiratory system, Rheumatoid arthritis (RA), CNS disorders, Down syndrome, Schizophrenia, nutritional deficiencies, primary open-angle glaucoma (POAG), and bone disorders, hematopoietic disorders, and/or other diseases or disorders.

In yet another aspect, the invention can be used in a method to identity the cellular receptors and downstream effectors of the invention by any one of a number of techniques commonly employed in the art. These include but are not limited to the two-hybrid system, affinity purification, co-precipitation with antibodies or other specific-interacting molecules.

Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Although methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, suitable methods and materials are described below. All publications, patent applications, patents, and other references mentioned herein are incorporated by reference in their entirety. In the case of conflict, the present specification, including definitions, will control. In addition, the materials, methods, and examples are illustrative only and not intended to be limiting.

Other features and advantages of the invention will be apparent from the following detailed description and claims.

DETAILED DESCRIPTION OF THE INVENTION

The invention is based, in part, upon the discovery of novel nucleic acid sequences that encode novel polypeptides. The novel nucleic acids and their encoded polypeptides are referred to individually as MOL1, MOL2, MOL3, MOL4, MOL5, MOL6, MOL7, and MOL8. The nucleic acids, and their encoded polypeptides, are collectively designated herein as "MOLX".

The novel MOLX nucleic acids of the invention include the nucleic acids whose sequences are provided in Tables 1A, 1D, 2A, 3A, 3D, 4A, 5A, 5C, 5E, 6A, 7A, and 8A. inclusive ("Tables 1A - 8A"), or a fragment, derivative, analog or homolog thereof. The novel MOLX proteins of the invention include the protein fragments whose sequences are provided in Tables 1B, 1E, 2B, 3B, 3E, 4B, 5B, 5D, 5F, 6B, 7B, and 8B inclusive ("Tables 1B - 8B"). The individual MOLX nucleic acids and proteins are described below. Within the scope of this invention is a method of using these nucleic acids and peptides in the treatment or prevention of a disorder related to cell signaling or metabolic pathway modulation.

MOL1

MOL1a

A disclosed Notch-like nucleic acid of 7410 nucleotides, MOL1a, alternatively referred to as SC29674552_EXT, is shown in Table 1A. The disclosed MOL1a open reading frame ("ORF") begins at the ATG initiation codon at nucleotides 1-3 and terminates at a TGA codon at nucleotides 7408-7410. In Table 1A, the start and stop codons are depicted with bold letters.

Table 1A. MOL1a nucleotide sequence (SEQ ID NO:1).

```
ATGCCCGCCCTGCGCCCCGCTCTGCTGTGGGCGCTGCTGGCGCTCTGGCTGTGCTGCGCGACCCCCGCGCATGCATT
GCAGTGTCGAGATGGCTATGAACCCTGTGTAAATGAAGGAATGTGTGTTACCTACCACAATGGCACAGGATACTGCA
AATGTCCAGAAGGCTTCTTGGGGGAATATTGTCAACATCGAGACCCCTGTGAGAAGAACCGCTGCCAGAATGGTGGG
ACTTGTGTGGCCCAGGCCATGCTGGGGAAAGCCACGTGCCGATGTGCCTCAGGGTTTACAGGAGAGGACTGCCAGTA
CTCGACATCTCATCCATGCTTTGTGTCTCGACCTTGCCTGAATGGCGGCACATGCCATATGCTCAGCCGGGATACCT
ATGAGTGCACCTGTCAAGTCGGGTTTACAGGTAAGGAGTGCCAATGGACGGATGCCTGCCTGTCTCATCCCTGTGCA
AATGGAAGTACCTGTACCACTGTGGCCAACCAGTTCTCCTGCAAATGCCTCACAGGCTTCACAGGGCAGAAATGTGA
GACTGATGTCAATGAGTGTGACATTCCAGGACACTGCCAGCATGGTGGCACCTGCCTCAACCTGCCTGGTTCCTACC
AGTGCCAGTGCCCTCAGGGCTTCACAGGCCAGTACTGTGACAGCCTGTATGTGCCCTGTGCACCCTCACCTTGTGTC
AATGGAGGCACCTGTCGGCAGACTGGTGACTTCACTTTTGAGTGCCATTTACCAGGTTTTGAAGGGAGCACCTGTGA
GAGGAATATTGATGACTGCCCTAACCACAGGTGTCAGAATGGAGGGGTTTGTGTGGATGGGGTCAACACTTACAACT
GCCGCTGTCCCCCACAATGGACAGGACAGTTCTGCACAGAGGATGTGGATGAATGCCTGCTGCAGCCCAATGCCTGT
CAAAACTGGGGCACCTGTGCCAACCGCAATGGAGGCTATGGCTGTGTATGTGTCAACGGCTGGAGTGGAGATGACTG
CAGTGAGAACATTGATGATTGTGCTTTCGGCGCCTGTACTCCAGGCTCCACCTGCATCGACCGTGTGGCCTCCTTCT
CTTGCATGTGCCCAGAGGGGAAGGCAGGTCTCCTGTGTCATCTGGATGATGCATGCATCAGCAATCCTTGCCACAAG
GGGGCACTGTGTGACACCAACCCCCTAAATGGGCAATATATTTGCACCTGCCCACAAGGCTACAAAGGGGCTGACTG
CACAGAAGATGTGGATGAATGTGCCATGGCCAATAGCAATCCTTGTGAGCATGCAGGAAAATGTGTGAACACGGATG
GCGCCTTCCACTGTGAGTGTCTGAAGGGTTATGCAGGACCTCGTTGTGAGATGGACATCAATGAGTGCCATTCAGAC
CCCTGCCAGAATGATGCTACCTGTCTGGATAAGATTGGAGGCTTCACATGTCTGTGCATGCCAGGTTTCAAAGGTGT
GCATTGTGAATTAGAAATAAATGAATGTCAGAGCAACCCTTGTGTGAACAATGGGCAGTGTGTGGATAAAGTCAATC
GTTTCCAGTGCCTGTGTCCTCCTGGTTTCACTGGGCCAGTTTGCCAGATTGATATTGATGACTGTTCCAGTACTCCG
```

TGTCTGAATGGGGCAAAGTGTATCGATCACCCGAATGGCTATGAATGCCAGTGTGCCACAGGTTTCACTGGTGTGTT
GTGTGAGGAGAACATTGACAACTGTGACCCCGATCCTTGCCACCATGGTCAGTGTCAGGATGGTATTGATTCCTACA
CCTGCATCTGCAATCCCGGGTACATGGGCGCCATCTGCAGTGACCAGATTGATGAATGTTACAGCAGCCCTTGCCTG
AACGATGGTCGCTGCATTGACCTGGTCAATGGCTACCAGTGCAACTGCCAGCCAGGCACGTCAGGTGTTAATTGTGA
AATTAATTTTGATGACTGTGCAAGTAACCCTTGTATCCATGGAATCTGTATGGATGGCATTAATCGCTACAGTTGTG
TCTGCTCACCAGGATTCACAGGGCAGAGATGTAACATTGACATTGATGAGTGTGCCTCCAATCCCTGTCGCAAGGGT
GCAACATGTATCAACGGTGTGAATGGTTTCCGCTGTATATGCCCCGAGGGACCCCATCACCCCAGCTGCTACTCACA
GGTGAACGAATGCCTGAGCAATCCCTGCATCCATGGAAACTGTACTGGAGGTCTCAGTGGATATAAGTGTCTCTGTG
ATGCAGGCTGGGTTGGCATCAACTGTGAAGTGGACAAAAATGAATGCCTTTCGAATCCATGCCAGAATGGAGGAACT
TGTGACAATCTGGTGAATGGATACAGGTGTACTTGCAAGAAGGGCTTTAAAGGCTATAACTGCCAGGTGAATATTGA
TGAATGTGCCTCAAATCCATGCCTGAACCAAGGAACCTGCTTTGATGACATAAGTGGCTACACTTGCCACTGTGTGC
TGCCATACACAGGTAAGAATTGTCAGACAGTATTGGCTCCCTGTTCCCCAAACCCTTGTGAGAATGCTGCTGTTTGC
AAAGAGTCACCAAATTTTGAGAGTTATACTTGCTTGTGTGCTCCTGGCTGGCAAGGTCAGCGGTGTACCATTGACAT
TGACGAGTGTATCTCCAAGCCCTGCATGAACCATGGTCTCTGCCATAACACCCAGGGCAGCTACATGTGTGAATGTC
CACCAGGCTTCAGTGGTATGGACTGTGAGGAGGACATTGATGACTGCCTTGCCAGTCCTTGCCAGAATGGAGGTTCC
TGTATGGATGGAGTGAATACTTTCTCCTGCCTCTGCCTTCCGGGTTTCACTGGGGATAAGTGCCAGACAGACATGAA
TGAGTGTCTGAGTGAACCCTGTAAGAATGGAGGGACCTGCTCTGACTACGTCAACAGTTACACTTGCAAGTGCCAGG
CAGGATTTGATGGAGTCCATTGTGAGAACAACATCAATGAGTGCACTGAGAGCTCCTGTTTCAATGGTGGCACATGT
GTTGATGGGATTAACTCCTTCTCTTGCTTGTGCCCTGTGGGTTTCACTGGATCCTTCTGCCTCCATGAGATCAATGA
ATGCAGCTCTCATCCATGCCTGAATGATGGAACGTGTGTTGATGGCCTGGGTACCTACCGCTGCAGCTGCCCCCTGG
GCTACACTGGGAAAAACTGTCAGACCCTGGTGAATCTCTGCAGTCGGTCTCCATGTAAAAACAAAGGTACTTGCGTT
CAGAAAAAAGCAGAGTCCCAGTGCCTATGTCCATCTGGATGGGCTGGTGCCTATTGTGACGTGCCCAATGTCTCTTG
TGAACATAGCAGCCTCCAGGAGAGGTGTGCTTGTTGAAACACTTGTGCCAGCACTCAGGTGTCTGCATCAATGCTGGCA
ACACGCATTACTGTCAGTGCCCCCTGGGCTATACTGGGAGCTACTGTGAGGAGCAACTCGATGAGTGTGCGTCCAAC
CCCTGCCAGCACGGGGCAACATGCAGTGACTTCATTGGTGGATACAGATGCGAGTGTGTCCCAGGCTATCAGGGTGT
CAACTGTGAGTATGAAGTGGATGAGTGCCAGAATCAGCCCTGCCAGAATGGAGGCACCTGTATTGACCTTGTGAACC
ATTTCAAGTGCTCTTGCCCACCAGGCACTCGGGGTATGAAATCATCCTTATCCATTTTCCATTGCCCGGGTCCCCAT
TGCCTTAATGGTGGTCAGTGCATGGATAGGATTGGAGGCTACAGTTGTCGCTGCTTGCCTGGCTTTGCTGGGGAGCG
TTGTGAGGGAGACATCAACGAGTGCCTCTCCAACCCCTGCAGCTCTGAGGGCAGCCTGGACTGTATACAGCTCACCA
ATGACTACCTGTGTGTTTGCCGTAGTGCCTTTACTGGTCGGCACTGTGAAACCTTCGTCGATGTGTGTCCCCAGATG
CCCTGCCTGAATGGAGGGACTTGTGCTGTGGCCAGTAACATGCCTGATGGTTCATTTGCCGTTGTCCCCCAGGGATT
TTCCGGGGCAAGGTGCCAGAGCAGCTGTGGACAAGTGAAATGTAGGAAGGGGGAGCAGTGTGTGCACACCGCCTCTG
GACCCCGCTGCTTCTGCCCCAGTCCCCGGGACTGCGAGTCAGGCTGTGCCAGTAGCCCCTGCCAGCACGGGGGCAGC
TGCCACCCTCAGCGCCAGCCTCCTTATTACTCCTGCCAGTGTGCCCCACCATTCTCGGGTAGCCGCTGTGAACTCTA
CACGGCACCCCCCAGCACCCCTCCTGCCACCTGTCTGAGCCAGTATTGTGCCGACAAAGCTCGGGATGGCGTCTGTG
ATGAGGCCTGCAACAGCCATGCCTGCCAGTGGGATGGGGGTGACTGTTCTCTCACCATGGAGAACCCCTGGGCCAAC
TGCTCCTCCCCACTTCCCTGCTGGGATTATATCAACAACCAGTGTGATGAGCTGTGCAACACGGTCGAGTGCCTGTT
TGACAACTTTGAATGCCAGGGGAACAGCAAGACATGCAAGTATGACAAATACTGTGCAGACCACTTCAAAGACAACC
ACTGTGACCAGGGGTGCAACAGTGAGGAGTGTGGTTGGGATGGGCTGGACTGTGCTGCTGACCAACCTGAGAACCTG
GCAGAAGGTACCCTGGTTATTGTGGTATTGATGCCACCTGAACAACTGCTCCAGGATGCTCGCAGCTTCTTGCGGGC
ACTGGGTACCCTGCTCCACACCAACCTGCGCATTAAGCGGGACTCCCAGGGGGAACTCATGGTGTACCCCTATTATG
GTGAGAAGTCAGCTGCTATGAAGAAACAGAGGATGACACGCAGATCCCTTCCTGGTGAACAAGAACAGGAGGTGGCT
GGGTCTAAAGTCTTTCTGGAAATTGACAACCGCCAGTGTGTTCAAGACTCAGACCACTGCTTCAAGAACACGGATGC
AGCAGCAGCTCTCCTGGCCTCTCACGCCATACAGGGGACCCTGTCATACCCTCTTGTGTCTGTCGTCAGTGAGTCCC
TGACTCCAGAACGCACTCAGCTCCTCTATCTCCTTGCTGTTGCTGTTGTCATCATTCTGTTTATTATTCTGCTGGGG
GTAATCATGGCAAAACGAAAGCGTAAGCATGGCTCTCTCTGGCTGCCTGAAGGTTTCACTCTTCGCCGAGATGCAAG
CAATCACAAGCGTCGTGAGCCAGTGGGACAGGATGCTGTGGGGCTGAAAAATCTCTCAGTGCAAGTCTCAGAAGCTA
ACCTAATTGGTACTGGAACAAGTGAACACTGGGTCGATGATGAAGGGCCCCAGCCAAAGAAAGTAAAGGCTGAAGAT
GAGGCCTTACTCTCAGAAGAAGATGACCCCATTGATCGACGGCCATGGACACAGCAGCACCTTGAAGCTGCAGACAT
CCGTAGGACACCATCGCTGGCTCTCACCCCTCCTCAGGCAGAGCAGGAGGTGGATGTGTTAGATGTGAATGTCCGTG
GCCCAGATGGCTGCACCCCATTGATGTTGGCTTCTCTCCGAGGAGGCAGCTCAGATTTGAGTGATGAAGATGAAGAT
GCAGAGGACTCTTCTGCTAACATCATCACAGACTTGGTCTACCAGGGTGCCAGCCTCCAGGCCCAGACAGACCGGAC
TGGTGAGATGGCCCTGCACCTTGCAGCCCGCTACTCACGGGCTGATGCTGCCAAGCGTCTCCTGGATGCAGGTGCAG
ATGCCAATGCCCAGGACAACATGGGCCGCTGTCCACTCCATGCTGCAGTGGCAGCTGATGCCCAAGGTGTCTTCCAG
ATTCTGATTCGCAACCGAGTAACTGATCTAGATGCCAGGATGAATGATGGTACTACACCCCTGATCCTGGCTGCCCG
CCTGGCTGTGGAGGGAATGGTGGCAGAACTGATCAACTGCCAAGCGGATGTGAATGCAGTGGATGACCATGGAAAAT
CTGCTCTTCACTGGGCAGCTGCTGTCAATAATGTGGAGGCAACTCTTTTGTTGTTGAAAAATGGGGCCAACCGAGAC
ATGCAGGACAACAAGGAAGAGACACCTCTGTTTCTTGCTGCCCGGGAGGGGAGCTATGAAGCAGCCAAGATCCTGTT
AGACCATTTTGCCAATCGAGACATCACAGACCATATGGATCGTCTTCCCCGGGATGTGGCTCGGGATCGCATGCACC
ATGACATTGTGCGCCTTCTGGATGAATACAATGTGACCCCAAGCCCTCCAGGCACCGTGTTGACTTCTGCTCTCTCA
CCTGTCATCTGTGGGCCCAACAGATCTTTCCTCAGCCTGAAGCACACCCCAATGGGCAAGAAGTCTAGACGGCCCAG
TGCCAAGAGTACCATGCCTACTAGCCTCCCTAACCTTGCCAAGGAGGCAAAGGATGCCAAGGGTAGTAGGAGGAAGA
AGTCTCTGAGTGAGAAGGTCCAACTGTCTGAGAGTTCAGTAACTTTATCCCCTGTTGATTCCCTAGAATCTCCTCAC
ACGTATGTTTCCGACACCACATCCTCTCCAATGATTACATCCCCTGGGATCTTACAGGCCTCACCCAACCCTATGTT
GGCCACTGCCGCCCCTCCTGCCCCAGTCCATGCCCAGCATGCACTATCTTTTTCTAACCTTCATGAAATGCAGCCTT
TGGCACATGGGGCCAGCACTGTGCTTCCCTCAGTGAGCCAGTTGCTATCCCACCACCACATTGTGTCTCCAGGCAGT
GGCAGTGCTGGAAGCTTGAGTAGGCTCCATCCAGTCCCAGTCCCAGCAGATTGGATGAACCGCATGGAGGTGAATGA
GACCCAGTACAATGAGATGTTTGGTATGGTCCTGGCTCCAGCTGTAGGGCACCCATCCTGGCATAGCTCCCCAGAGA
GGCCACCTGAAGGGAAGCACATAACCACCCCTCGGGAGCCCTTGCCCCCCCATTGTGACTTTCCAGCTCATCCCTAAA

GGCAGTATTGCCCAACCAGCGGGGGCTCCCCAGCCTCAGTCCACCTGCCCTCCAGCTGTTGCGGGCCCCCTGCCCAC
CATGTACCAGATTCCAGAAATGGCCCGTTTGCCCAGTGTGGCTTTCCCCACTGCCATGATGCCCCAGCAGGACGGGC
AGGTAGCTCAGACCATTCTCCCAGCCTATCATCCTTTCCCAGCCTCTGTGGGCAAGTACCCCACACCCCCTTCACAG
CACAGTTATGCTTCCTCAAATGCTGCTGAGCGAACACCCAGTCACAGTGGTCACCTCCAGGGTGAGCATCCCTACCT
GACACCATCCCCAGAGTCTCCTGACCAGTGGTCAAGTTCATCACCCCACTCTGCTTCTGACTGGTCAGATGTGACCA
CCAGCCCTACCCCTGGGGGAGCTGGAGGAGGTCAGCGGGGACCTGGGACACACATGTCTGAGCCACCACACAACAAC
ATGCAGGTTTATGCGTGA

The disclosed MOL1a nucleotide encodes a protein which has 2469 amino acid residues, referred to as the MOL1a protein. The MOL1a protein was analyzed for signal peptide prediction and cellular localization. SignalP results predict that MOL1a is cleaved between position 25 and 26 (AHA-LQ) of SEQ ID NO:2. Psort and Hydropathy profiles also predict that MOL1a contains a signal peptide and is likely to be localized in the plasma membrane (Certainty=0.4600). A disclosed MOL1a polypeptide sequence is presented in Table 1B using the one-letter amino acid code.

Table 1B. Encoded MOL1a protein sequence (SEQ ID NO:2).

MPALRPALLWALLALWLCCATPAHALQCRDGYEPCVNEGMCVTYHNGTGYCKCPEGFLGEYCQHRDPCEKNRCQNGG
TCVAQAMLGKATCRCASGFTGEDCQYSTSHPCFVSRPCLNGGTCHMLSRDTYECTCQVGFTGKECQWTDACLSHPCA
NGSTCTTVANQFSCKCLTGFTGQKCETDVNECDIPGHCQHGGTCLNLPGSYQCQCPQGFTGQYCDSLYVPCAPSPCV
NGGTCRQTGDFTFECHLPGFEGSTCERNIDDCPNHRCQNGGVCVDGVNTYNCRCPPQWTGQFCTEDVDECLLQPNAC
QNWGTCANRNGGYGCVCVNGWSGDDCSENIDDCAFGACTPGSTCIDRVASFSCMCPEGKAGLLCHLDDACISNPCHK
GALCDTNPLNGQYICTCPQGYKGADCTEDVDECAMANSNPCEHAGKCVNTDGAFHCECLKGYAGPRCEMDINECHSD
PCQNDATCLDKIGGFTCLCMPGFKGVHCELEINECQSNPCVNNGQCVDKVNRFQCLCPPGFTGPVCQIDIDDCSSTP
CLNGAKCIDHPNGYECQCATGFTGVLCEENIDNCDPDPCHHGQCQDGIDSYTCICNPGYMGAICSDQIDECYSSPCL
NDGRCIDLVNGYQCNCQPGTSGVNCEINFDDCASNPCIHGICMDGINRYSCVCSPGFTGQRCNIDIDECASNPCRKG
ATCINGVNGFRCICPEGPHHPSCYSQVNECLSNPCIHGNCTGGLSGYKCLCDAGWVGINCEVDKNECLSNPCQNGGT
CDNLVNGYRCTCKKGFKGYNCQVNIDECASNPCLNQGTCFDDISGYTCHCVLPYTGKNCQTVLAPCSPNPCENAAVC
KESPNFESYTCLCAPGWQGQRCTIDIDECISKPCMNHGLCHNTQGSYMCECPPGFSGMDCEEDIDDCLASPCQNGGS
CMDGVNTFSCLCLPGFTGDKCQTDMNECLSEPCKNGGTCSDYVNSYTCKCQAGFDGVHCENNINECTESSCFNGGTC
VDGINSFSCLCPVGFTGSFCLHEINECSSHPCLNDGTCVDGLGTYRCSCPLGYTGKNCQTLVNLCSRSPCKNKGTCV
QKKAESQCLCPSGWAGAYCDVPNVSCDIAASRRGVLVEHLCQHSGVCINAGNTHYCQCPLGYTGSYCEEQLDECASN
PCQHGATCSDFIGGYRCECVPGYQGVNCEYEVDECQNQPCQNGGTCIDLVNHFKCSCPPGTRGMKSSLSIFHCPGPH
CLNGGQCMDRIGGYSCRCLPGFAGERCEGDINECLSNPCSSEGSLDCIQLTNDYLCVCRSAFTGRHCETFVDVCPQM
PCLNGGTCAVASNMPDGSFAVVPQGFSGARCQSSCGQVKCRKGEQCVHTASGPRCFCPSPRDCESGCASSPCQHGGS
CHPQRQPPYYSCQCAPPFSGSRCELYTAPPSTPPATCLSQYCADKARDGVCDEACNSHACQWDGGDCSLTMENPWAN
CSSPLPCWDYINNQCDELCNTVECLFDNFECQGNSKTCKYDKYCADHFKDNHCDQGCNSEECGWDGLDCAADQPENL
AEGTLVIVVLMPPEQLLQDARSFLRALGTLLHTNLRIKRDSQGELMVYPYYGEKSAAMKKQRMTRRSLPGEQEQEVA
GSKVFLEIDNRQCVQDSDHCFKNTDAAAALLASHAIQGTLSYPLVSVVSESLTPERTQLLYLLAVAVVIILFIILLG
VIMAKRKRKHGSLWLPEGFTLRRDASNHKRREPVGQDAVGLKNLSVQVSEANLIGTGTSEHWVDDEGPQPKKVKAED
EALLSEEDDPIDRRPWTQQHLEAADIRRTPSLALTPPQAEQEVDVLDVNVRGPDGCTPLMLASLRGGSSDLSDEDED
AEDSSANIITDLVYQGASLQAQTDRTGEMALHLAARYSRADAAKRLLDAGADANAQDNMGRCPLHAAVAADAQGVFQ
ILIRNRVTDLDARMNDGTTPLILAARLAVEGMVAELINCQADVNAVDDHGKSALHWAAAVNNVEATLLLLKNGANRD
MQDNKEETPLFLAARFGSYEAAKILLDHFANRDITDHMDRLPRDVARDRMHHDIVRLLDEYNVTPSPPGTVLTSALS
PVICGPNRSFLSLKHTPMGKKSRRPSAKSTMPTSLPNLAKEAKDAKGSRRKKSLSEKVQLSESSVTLSPVDSLESPH
TYVSDTTSSPMITSPGILQASPNPMLATAAPPAPVHAQHALSFSNLHEMQPLAHGASTVLPSVSQLLSHHHIVSPGS
GSAGSLSRLHPVPVPADWMNRMEVNETQYNEMPGMVLAPAVGHPSWHSSPERPPEGKHITTPREPLPPIVTFQLIPK
GSIAQPAGAPQPQSTCPPAVAGPLPTMYQIPEMARLPSVAFPTAMMPQQDGQVAQTILPAYHPFPASVGKYPTPPSQ
HSYASSNAAERTPSHSGHLQGEHPYLTPSPESPDQWSSSSPHSASDWSDVTTSPTPGGAGGGQRGPGTHMSEPPHNN
MQVYA

A region of the MOL1a nucleic acid sequence has 6436 of 7416 bases (86%) identical to a *Rattus norvegicus* Notch-like protein mRNA (GENBANK-ID:RATNOTCHX | acc:M93661), with an E-value of 0.0. In all BLAST alignments herein, the "E-value" or

"Expect" value is a numeric indication of the probability that the aligned sequences could have achieved their similarity to the BLAST query sequence by chance alone, within the database that was searched. For example, the probability that the subject ("Sbjct") retrieved from the MOL1a BLAST analysis, *e.g.*, the *Rattus norvegicus* Notch-like protein mRNA, matched the Query MOL1a sequence purely by chance is 0.0. MOL1a also has 2443 of 2471 amino acid residues (98%) positive with patp:AAY06816 Human Notch2 (humN2) protein sequence - *Homo sapiens*, 2471 aa.

The Expect value is used as a convenient way to create a significance threshold for reporting results. The default value used for blasting is typically set to 0.0001. In BLAST 2.0, the Expect value is also used instead of the P value (probability) to report the significance of matches. For example, an E value of one assigned to a hit can be interpreted as meaning that in a database of the current size one might expect to see one match with a similar score simply by chance. An E value of zero means that one would not expect to see any matches with a similar score simply by chance. See, *e.g.*, http://www.ncbi.nlm.nih.gov/Education/BLASTinfo/. Occasionally, a string of X's or N's will result from a BLAST search. This is a result of automatic filtering of the query for low-complexity sequence that is performed to prevent artifactual hits. The filter substitutes any low-complexity sequence that it finds with the letter "N" in nucleotide sequence (*e.g.*, "NNNNNNNNNNNNN") or the letter "X" in protein sequences (*e.g.*, "XXXXXXXXX"). Low-complexity regions can result in high scores that reflect compositional bias rather than significant position-by-position alignment. Wootton and Federhen, Methods Enzymol 266:554-571, 1996.

Utilities for the MOLX nucleic acids and their encoded polypeptides can be inferred based on the homology of the disclosed MOLX nucleic acids and/or polypeptides (including domains of the encoded polypeptides) to previously described sequences.

MOL1a expression in different tissues was examined through TaqMan as described below in Example 1.

MOL1a is expressed in at least the following tissues: kidney, brain, lymph node, muscle, hippocampus, bone marrow, placenta, thyroid, para-thyroid, prostate, testis, epidermis, ovary, coronary artery, liver, lung, spinal cord, stomach, breast, lung, uterus, and colon. It is likely that Notch proteins are expressed in all tissues, so the widespread expression of MOL1a agrees with its homology with Notch.

One or more consensus positions (Cons. Pos.) of the nucleotide sequence of MOL1a have been identified as single nucleotide polymorphisms (SNPs) as shown in Table 1C. A

dash ("-"), when shown, means that a base is not present. The sign ">" means "is changed to". SNPs were identified using the techniques disclosed in Example 3.

Table 1C: SNPs for MOL1a			
Consensus Position	**Base Change**	**AA change Position**	**Residue Change**
4288	G > A	1429	A > T
5858	T > C	1952	L > P
5833	A > G	1944	T > A
5366	C > T	1788	T > I

MOL1b

MOL1a was subjected to the exon linking process to confirm the sequence. PCR primers were designed by starting at the most upstream sequence available, for the forward primer, and at the most downstream sequence available for the reverse primer. In each case, the sequence was examined, walking inward from the respective termini toward the coding sequence, until a suitable sequence that is either unique or highly selective was encountered, or, in the case of the reverse primer, until the stop codon was reached. Such primers were designed based on *in silico* predictions for the full length cDNA, part (one or more exons) of the DNA or protein sequence of the target sequence, or by translated homology of the predicted exons to closely related human sequences sequences from other species. These primers were then employed in PCR amplification based on the following pool of human cDNAs: adrenal gland, bone marrow, brain - amygdala, brain - cerebellum, brain - hippocampus, brain - substantia nigra, brain - thalamus, brain -whole, fetal brain, fetal kidney, fetal liver, fetal lung, heart, kidney, lymphoma - Raji, mammary gland, pancreas, pituitary gland, placenta, prostate, salivary gland, skeletal muscle, small intestine, spinal cord, spleen, stomach, testis, thyroid, trachea, uterus. Usually the resulting amplicons were gel purified, cloned and sequenced to high redundancy. The resulting sequences from all clones were assembled with themselves, with other fragments in CuraGen Corporation's database and with public ESTs. Fragments and ESTs were included as components for an assembly when the extent of their identity with another component of the assembly was at least 95% over 50 bp. In addition, sequence traces were evaluated manually and edited for corrections if appropriate. These procedures provide the sequence reported below, which is designated MOL1b, or alternatively Accession Number CG56250-02. This differs from the previously identified sequence in lacking 996 internal amino acids in addition to a few minor changes.

A disclosed Notch-like nucleic acid of 6728 nucleotides, MOL1b, is shown in Table 1C. The disclosed MOL1b open reading frame ("ORF") begins at the ATG initiation codon at nucleotides 13-15, and terminates at a TGA codon at nucleotides 4431-4434. In Table 1D the start and stop codons are in bold letters, and the untranslated regions are underlined.

5

Table 1D. MOL1b nucleotide sequence (SEQ ID NO:3).

TCATCTGGAATTATGCCCGCCCTGCGCCCCGCTCTGCTGTGGGCGCTGCTGGCGCTCTGGCTGTGCTGCGCGGCCCC
CGCGCATGCATTGCAGTGTCGAGATGGCTATGAACCCTGTGTAAATGAAGGAATGTGTGTTACCTACCACAATGGCA
CAGGATACTGCAAATGTCCAGAAGGCTTCTTGGGGGAATATTGTCAACATCGAGACCCCTGTGAGAAGAACCGCTGC
CAGAATGGTGGGACTTGTGTGGCCCAGGCCATGCTGGGGAAAGCCACGTGCCGATGTGCCTCAGGGTTTACAGGAGA
GGACTGCCAGTACTCAACATCTCATCCATGCTTTGTGTCTCGACCCTGCCTGAATGGCGGCACATGCCATATGCTCA
GCCGGGATACCTATGAGTGCACCTGTCAAGTCGGGTTTACAGGTAAGGAGTGCCAATGGACGGATGCCTGCCTGTCT
CATCCCTGTGCAAATGGAAGTACCTGTACCACTGTGGCCAACCAGTTCTCCTGCAAATGCCTCACAGGCTTCACAGG
GCAGAAATGTGAGACTGATGTCAATGAGTGTGACATTCCAGGACACTGCCAGCATGGTGGCACCTGCCTCAACCTGC
CTGGTTCCTACCAGTGCCAGTGCCCTCAGGGCTTCACAGGCCAGTACTGTGACAGCCTGTATGTGCCCTGTGCACCC
TCACCTTGTGTCAATGGAGGCACCTGTCGGCAGACTGGTGACTTCACTTTTGAGTGCAACTGCCTTCCAGGTTTTGA
AGGGAGCACCTGTGAGAGGAATATTGATGACTGCCCTAACCACAGGTGTCAGAATGGAGGGGTTTGTGTGGATGGGG
TCAACACTTACAACTGCCGCTGTCCCCCACAATGGACAGGACAGTTCTGCACAGAGGATGTGGATGAATGCCTGCTG
CAGCCCAATGCCTGTCAAAATGGGGGCACCTGTGCCAACCGCAATGGAGGCTATGGCTGTGTATGTGTCAACGGCTG
GAGTGGAGATGACTGCAGTGAGAACATTGATGATTGTGCCTTCGCCTCCTGTACTCCAGGCTCCACCTGCATCGACC
GTGTGGCCTCCTTCTCTTGCATGTGCCCAGAGGGGAAGGCAGGTCTCCTGTGTCATCTGGATGATGCATGCATCAGC
AATCCTTGCCACAAGGGGGCACTGTGTGACACCAACCCCCTAAATGGGCAATATATTTGCACCTGCCCACAAGGCTA
CAAAGGGGCTGACTGCACAGAAGATGTGGATGAATGTGCCATGGCCAATAGCAATCCTTGTGAGCATGCAGGAAAAT
GTGTGAACACGGATGGCGCCTTCCACTGTGAGTGTCTGAAGGGTTATGCAGGACCTCGTTGTGAGATGGACATCAAT
GAGTGCCATTCAGACCCCTGCCAGAATGATGCTACCTGTCTGGATAAGATTGGAGGCTTCACATGTCTGTGCATGCC
AGGTTTCAAAGGTGTGCATTGTGAATTAGAAATAAATGAATGTCAGAGCAACCCTTGTGTGAACAATGGGCAGTGTG
TGGATAAAGTCAATCGTTTCCAGTGCCTGTGTCCTCCTGGTTTCACTGGGCCAGTTTGCCAGATTGATATTGATGAC
TGTTCCAGTACTCCGTGTCTGAATGGGGCAAAGTGTATCGATCACCCGAATGGCTATGAATGCCAGTGTGCCACAGG
TTTCACTGGTGTGTTGTGTGAGGAGAACATTGACAACTGTGACCCCGATCCTTGCCACCATGGTCAGTGTCAGGATG
GTATTGATTCCTACACCTGCATCTGCAATCCCGGGTACATGGGCGCCATCTGCAGTGACCAGATTGATGAATGTTAC
AGCAGCCCTTGCCTGAACGATGGTCGCTGCATTGACCTGGTCAATGGCTACCAGTGCAACTGCCAGCCAGGCACGTC
AGGGGTTAATTGTGAAATTAATTTTGATGACTGTGCAAGTAACCCTTGTATCCATGGAATCTGTATGGATGGCATTA
ATCGCTACAGTTGTGTCTGCTCACCAGGATTCACAGGGCAGAGATGTAACATTGACATTGATGAGTGTGCCTCCAAT
CCCTGTCGCAAGGGTGCAACATGTATCAACGGTGTGAATGGTTTCCGCTGTATATGCCCCGAGGGACCCCATCACCC
CAGCTGCTACTCACAGGTGAACGAATGCCTGAGCAATCCCTGCATCCATGGAAACTGTACTGGAGGTCTCAGTGGAT
ATAAGTGTCTCTGTGATGCAGGCTGGGTTGGCATCAACTGTGAAGTGGACAAAAATGAATGCCTTTCGAATCCATGC
CAGAATGGAGGAACTTGTGACAATCTGGTGAATGGATACAGGTGTACTTGCAAGAAGGGCTTTAAAGGCTATAACTG
CCAGGTGAATATTGATGAATGTGCCTCAAATCCATGCCTGAACCAAGGAACCTGCTTTGATGACATAAGTGGCTACA
CTTGCCACTGTGTGCTGCCATACACAGGCAAGAATTGTCAGACAGTATTGGCTCCCTGTTCCCCAAACCCTTGTGAG
AATGCTGCTGTTTGCAAAGAGTCACCAAATTTTGAGAGTTATACTTGCTTGTGTGCTCCTGGCTGGCAAGGTCAGCG
GTGTACCATTGACATTGACGAGTGTATCTCCAAGCCCTGCATGAACCATGGTCTCTGCCATAACACCCAGGGCAGCT
ACATGTGTGAATGTCCACCAGGCTTCAGTGGTATGGACTGTGAGGAGGACATTGATGACTGCCTTGCCAATCCTTGC
CAGAATGGAGGTTCCTGTATGGATGGAGTGAATACTTTCTCCTGCCTCTGCCTTCCGGGTTTCACTGGGGATAAGTG
CCAGACAGACATGAATGAGTGTCTGAGTGAACCCTGTAAGAATGGAGGGACCTGCTCTGACTACGTCAACAGTTACA
CTTGCAAGTGCCAGGCAGGATTTGATGGAGTCCATTGTGAGAACAACATCAATGAGTGCACTGAGAGCTCCTGTTTC
AATGGTGGCACATGTGTTGATGGGATTAACTCCTTCTCTTGCTTGTGCCCTGTGGGTTTCACTGGATCCTTCTGCCT
CCATGAGATCAATGAATGCAGCTCTCATCCATGCCTGAATGAGGGAACGTGTGTTGATGGCCTGGGTACCTACCGCT
GCAGCTGCCCCCTGGGCTACACTGGGAAAAACTGTCAGACCCTGGTGAATCTCTGCAGTCGGTCTCCATGTAAAAAC
AAAGGTACTTGTGTTCAGAAAAAAGCAGAGTCCCAGTGCCTATGTCCATCTGGATGGGCTGGTGCCTATTGTGACGT
GCCCAATGTCTCTTGTGACATAGCAGCCTCCAGGAGAGGTGTGCTTGTTGAACACTTGTGCCAGCACTCAGGTGTCT
GCATCAATGCTGGCAACACGCATTACTGTCAGTGCCCCCTGGGCTATACTGGGAGCTACTGTGAGGAGCAACTCGAT
GAGTGTGCGTCCAACCCCTGCCAGCACGGGGCAACATGCAGTGACTTCATTGGTGGATACAGATGCGAGTGTGTCCC
AGGCTATCAGGGTGTCAACTGTGAGTATGAAGTGGATGAGTGCCAGAATCAGCCCTGCCAGAATGGAGGCACCTGTA
TTGACCTTGTGAACCATTTCAAGTGCTCTTGCCCACCAGGCACTCGGGGCCTACTCTGTGAAGAGAACATTGATGAC
TGTGCCCGGGGTCCCCATTGCCTTAATGGTGGTCAGTGCATGGATAGGATTGGAGGCTACAGTTGTCGCTGCTTGCC
TGGCTTTGCTGGGGAGCGTTGTGAGGGAGACATCAACGAGTGCCTCTCCAACCCCTGCAGCTCTGAGGGCAGCCTGG
ACTGTATACAGCTCACCAATGACTACCTGTGTGTTTGCCGTAGTGCCTTTACTGGCCGGCACTGTGAAACCTTCGTC
GATGTGTGTCCCCAGATGCCCTGCCTGAATGGAGGGACTTGTGCTGTGGCCAGTAACATGCCTGATGGTTTCATTTG
CCGTTGTCCCCCGGGATTTTCCGGGGCAAGGTACCAGATTCCAGAAATGGCCCGTTTGCCCAGTGTGGCTTTCCCCA
CTGCCATGATGCCCCAGCAGGACGGGCAGGTAGCTCAGACCATTCTCCCAGCCTATCATCCTTTCCCAGCCTCTGTG
GGCAAGTACCCCACACCCCCTTCACAGCACAGTTATGCTTCCTCAAATGCTGCTGAGCGAACACCCAGTCACAGTGG
TCACCTCCAGGGTGAGCATCCCTACCTGACACCATCCCCAGAGTCTCCTGACCAGTGGTCAAGTTCATCACCCCACT
CTGCTTCTGACTGGTCAGATGTGACCACCAGCCCTACCCCTGGGGGTGCTGGAGGAGGTCAGCGGGGACCTGGGACA

CACATGTCTGAGCCACCACACAACAACATGCAGGTTTATGCGTGAGAGAGTCCACCTCCAGTGTAGAGACATAACTG
ACTTTTGTAAATGCTGCTGAGGAACAAATGAAGGTCATCCGGGAGAGAAATGAAGAAATCTCTGGAGCCAGCTTCTA
GAGGTAGGAAAGAGAAGATGTTCTTATTCAGATAATGCAAGAGAAGCAATTCGTCAGTTTCACTGGGTATCTGCAAG
GCTTATTGATTATTCTAATCTAATAAGACAAGTTTGTGGAAATGCAAGATGAATACAAGCCTTGGGTCCATGTTTAC
TCTCTTCTATTTGGAGAATAAGATGGATGCTTATTGAAGCCCAGACATTCTTGCAGCTTGGACTGCATTTTAAGCCC
TGCAGGCTTCTGCCATATCCATGAGAAGATTCTACACTAGCGTCCTGTTGGGAATTATGCCCTGGAATTCTGCCTGA
ATTGACCTACGCATCTCCTCCTCCTTGGACATTCTTTTGTCTTCATTTGGTGCTTTTGGTTTTGCACCTCTCCGTGA
TTGTAGCCCTACCAGCATGTTATAGGGCAAGACCTTTGTGCTTTTGATCATTCTGGCCCATGAAAGCAACTTTGGTC
TCCTTTCCCCTCCTGTCTTCCCGGTATCCCTTGGAGTCTCACAAGGTTTACTTTGGTATGGTTCTCAGCACAAACCT
TTCAAGTATGTTGTTTCTTTGGAAAATGGACATACTGTATTGTGTTCTCCTGCATATATCATTCCTGGAGAGAGAAG
GGGAGAAGAATACTTTTCTTCAACAAATTTTGGGGGCAGGAGATCCCTTCAAGAGGCTGCACCTTAATTTTTCTTGT
CTGTGTGCAGGTCTTCATATAAACTTTACCAGGAAGAAGGGTGTGAGTTTGTTGTTTTTCTGTGTATGGGCCTGGTC
AGTGTAAAGTTTTATCCTTGATAGTCTAGTTACTATGACCCTCCCCACTTTTTTAAAACCAGAAAAAGGTTTGGAAT
GTTGGAATGACCAAGAGACAAGTTAACTCGTGCAAGAGCCAGTTACCCACCCACAGGTCCCCCTACTTCCTGCCAAG
CATTCCATTGACTGCCTGTATGGAACACATTTGTCCCAGATCTGAGCATTCTAGGCCTGTTTCACTCACTCACCCAG
CATATGAAACTAGTCTTAACTGTTGAGCCTTTCCTTTCATATCCACAGAAGACACTGTCTCAAATGTTGTACCCTTG
CCATTTAGGACTGAACTTTCCTTAGCCCAAGGGACCCAGTGACAGTTGTCTTCCGTTTGTCAGATGATCAGTCTCTA
CTGATTATCTTGCTGCTTAAAGGCCTGCTCACCAATCTTTCTTTCACACCGTGTGGTCCGTGTTACTGGTATACCCA
GTATGTTCTCACTGAAGACATGGACTTTATATGTTCAAGTGCAGGAATTGGAAAGTTGGACTTGTTTTCTATGATCC
AAAACAGCCCTATAAGAAGGTTGGAAAAGGAGGAACTATATAGCAGCCTTTGCTATTTTCTGCTACCATTTCTTTTC
CTCTGAAGCGGCCATGACATTCCCTTTGGCAACTAACGTAGAAACTCAACAGAACATTTTCCTTTCCTAGAGTCACC
TTTTAGATGATAATGGACAACTATAGACTTGCTCATTGTTCAGACTGATTGCCCCTCACCTGAATCCACTCTCTGTA
TTCATGCTCTTGGCAATTTCTTTGACTTTCTTTTAAGGGCAGAAGCATTTTAGTTAATTGTAGATAAAGAATAGTTT
TCTTCCTCTTCTCCTTGGGCCAGTTAATAATTGGTCCATGGCTACACTGCAACTTCCGTCCAGTGCTGTGATGCCCA
TGACACCTGCAAAATAAGTTCTGCCTGGGCATTTTGTAGATATTAACAGGTGAATTCCCGACTCTTTTGGTTTGAAT
GACAGTTCTCATTCCTTCTATGGCTGCAAGTATGCATCAGTGCTTCCCACTTACCTGATTTGTCTGTCGGTGGCCCC
ATATGGAAACCCTGCGTGTCTGTTGGCATAATAGTTTACAAATGGTTTTTTCAGTCCTATCCAAATTTATTGAACCA
ACAAAAATAATTACTTCTGCCCTGAGATAAGCAGATTAAGTTTGTTCATTCTCTGCTTTATTCTCTCCATGTGGCAA
CATTCTGTCAGCCTCTTTCATAGTGTGCAAACATTTTATCATTCTAAATGGTGACTCTCTGCCCTTGGACCCATTTA
TTATTCACAGATGGGGAGAACCTATCTGCATGGACCCTCACCATCCTCTGTGCAGCACACACAGTGCAGGGAGCCAG
TGGCGATGGCGATGACTTTCTTCCCCTGG

The protein encoded by the MOL1b nucleic acid sequence has 2469 amino acid residues, and is disclosed in Table 1E. The MOL1b protein was analyzed for signal peptide prediction and cellular localization. SignalP results predict that MOL1b is cleaved between position 25 and 26 (AHA-LQ) of SEQ ID NO:4. Psort and Hydropathy profiles also predict that MOL1b contains a signal peptide and is likely to be localized extracellularly (Certainty=0.7666).

Table 1E. Encoded MOL1b protein sequence (SEQ ID NO:4).

MPALRPALLWALLALWLCCAAPAHALQCRDGYEPCVNEGMCVTYHNGTGYCKCPEGFLGEYCQHRDPCEKNRCQNGG
TCVAQAMLGKATCRCASGFTGEDCQYSTSHPCFVSRPCLNGGTCHMLSRDTYECTCQVGFTGKECQWTDACLSHPCA
NGSTCTTVANQFSCKCLTGFTGQKCETDVNECDIPGHCQHGGTCLNLPGSYQCQCPQGFTGQYCDSLYVPCAPSPCV
NGGTCRQTGDFTFECNCLPGFEGSTCERNIDDCPNHRCQNGGVCVDGVNTYNCRCPPQWTGQFCTEDVDECLLQPNA
CQNGGTCANRNGGYGCVCVNGWSGDDCSENIDDCAFASCTPGSTCIDRVASFSCMCPEGKAGLLCHLDDACISNPCH
KGALCDTNPLNGQYICTCPQGYKGADCTEDVDECAMANSNPCEHAGKCVNTDGAFHCECLKGYAGPRCEMDINECHS
DPCQNDATCLDKIGGFTCLCMPGFKGVHCELEINECQSNPCVNNGQCVDKVNRFQCLCPPGFTGPVCQIDIDDCSST
PCLNGAKCIDHPNGYECQCATGFTGVLCEENIDNCDPDPCHHGQCQDGIDSYTCICNPGYMGAICSDQIDECYSSPC
LNDGRCIDLVNGYQCNCQPGTSGVNCEINFDDCASNPCIHGICMDGINRYSCVCSPGFTGQRCNIDIDECASNPCRK
GATCINGVNGFRCICPEGPHHPSCYSQVNECLSNPCIHGNCTGGLSGYKCLCDAGWVGINCEVDKNECLSNPCQNGG
TCDNLVNGYRCTCKKGFKGYNCQVNIDECASNPCLNQGTCFDDISGYTCHCVLPYTGKNCQTVLAPCSPNPCENAAV
CKESPNFESYTCLCAPGWQGQRCTIDIDECISKPCMNHGLCHNTQGSYMCECPPGFSGMDCEEDIDDCLANPCQNGG
SCMDGVNTFSCLCLPGFTGDKCQTDMNECLSEPCKNGGTCSDYVNSYTCKCQAGFDGVHCENNINECTESSCFNGGT
CVDGINSFSCLCPVGFTGSFCLHEINECSSHPCLNEGTCVDGLGTYRCSCPLGYTGKNCQTLVNLCSRSPCKNKGTC
VQKKAESQCLCPSGWAGAYCDVPNVSCDIAASRRGVLVEHLCQHSGVCINAGNTHYCQCPLGYTGSYCEEQLDECAS
NPCQHGATCSDFIGGYRCECVPGYQGVNCEYEVDECQNQPCQNGGTCIDLVNHFKCSCPPGTRGLLCEENIDDCARG
PHCLNGGQCMDRIGGYSCRCLPGFAGERCEGDINECLSNPCSSEGSLDCIQLTNDYLCVCRSAFTGRHCETFVDVCP
QMPCLNGGTCAVASNMPDGFICRCPPGFSGARYQIPEMARLPSVAFPTAMMPQQDGQVAQTILPAYHPFPASVGKYP
TPPSQHSYASSNAAERTPSHSGHLQGEHPYLTPSPESPDQWSSSSPHSASDWSDVTTSPTPGGAGGGQRGPGTHMSE

PPHNNMQVYA

A region of the MOL1b nucleic acid sequence, localized to chromosome 1, has 4041 of 4042 bases (99%) identical to a gb:GENBANK-ID:AF308601|acc:AF308601.1 mRNA from Homo sapiens (Homo sapiens NOTCH 2 (N2) mRNA, complete cds), with an E-value of 0.0.

The amino acid sequence of MOL1b has 1340 of 1343 amino acid residues (99%) identical to, and 1340 of 1343 amino acid residues (99%) similar to, the 2471 amino acid residue ptnr:TREMBLNEW-ACC:AAG37073 protein from Homo sapiens (Human) (NOTCH2 PROTEIN).

MOL1b expressed in at least the following tissues: adrenal gland, bone marrow, brain - amygdala, brain - cerebellum, brain - hippocampus, brain - substantia nigra, brain - thalamus, brain -whole, fetal brain, fetal kidney, fetal liver, fetal lung, heart, kidney, lymphoma - Raji, mammary gland, pancreas, pituitary gland, placenta, prostate, salivary gland, skeletal muscle, small intestine, spinal cord, spleen, stomach, testis, thyroid, trachea, uterus, Aorta, Ascending Colon, Bone, Cartilage, Cochlea, Colon, Coronary Artery, Epidermis, Foreskin, Liver, Lung, Lymph node, Lymphoid tissue, Muscle, Nasoepithelium, Ovary, Parathyroid Gland, Parotid Salivary glands, Peripheral Blood, Respiratory Bronchiole, Retina, Synovium/Synovial membrane, Thymus, Tonsils, Umbilical Vein, Vein, Whole Organism.

One or more consensus positions (Cons. Pos.) of the nucleotide sequence of MOL1b have been identified as single nucleotide polymorphisms (SNPs) as shown in Table 1F. "Depth" represents the number of clones covering the region of the SNP. The Putative Allele Frequency (PAF) is the fraction of all the clones containing the SNP. A dash ("-"), when shown, means that a base is not present. The sign ">" means "is changed to".

Table 1F: SNPs for MOL1b

Consensus Position	Depth	Base Change	PAF
100	24	T > C	0.083
204	23	G > A	0.087
369	23	G > A	0.087

The amino acid sequence of MOL1a also had high homology to other proteins as shown in table 1G.

Table 1G. BLAST results for MOL1a

Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi|104252|pir||A35844	Notch protein - African clawed frog	2524	1323/2515 (52%)	1676/2515 (66%),	0.0
>gi|6679096|ref|NP_032742.1|	Notch gene homolog 3, (Drosophila) [Mus musculus]	2318	665/1515 (43%)	860/1515 (55%)	0.0
gi|13242247|ref|NP_077334.1|	Notch gene homolog 2, (Drosophila) [Rattus norvegicus]	2471	2251/2472 (91%)	2340/2472 (94%)	0.0
gi|2209059|dbj|BAA20535.1|	Notch 2 [Takifugu rubripes]	2447	1240/2203 (56%)	1546/2203 (69%)	0.0
gi|6093542|sp|Q07008|NTC1_RAT	NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 1 PRECURSOR	2531	1340/2538 (52%)	1683/2538 (65%)	0.0

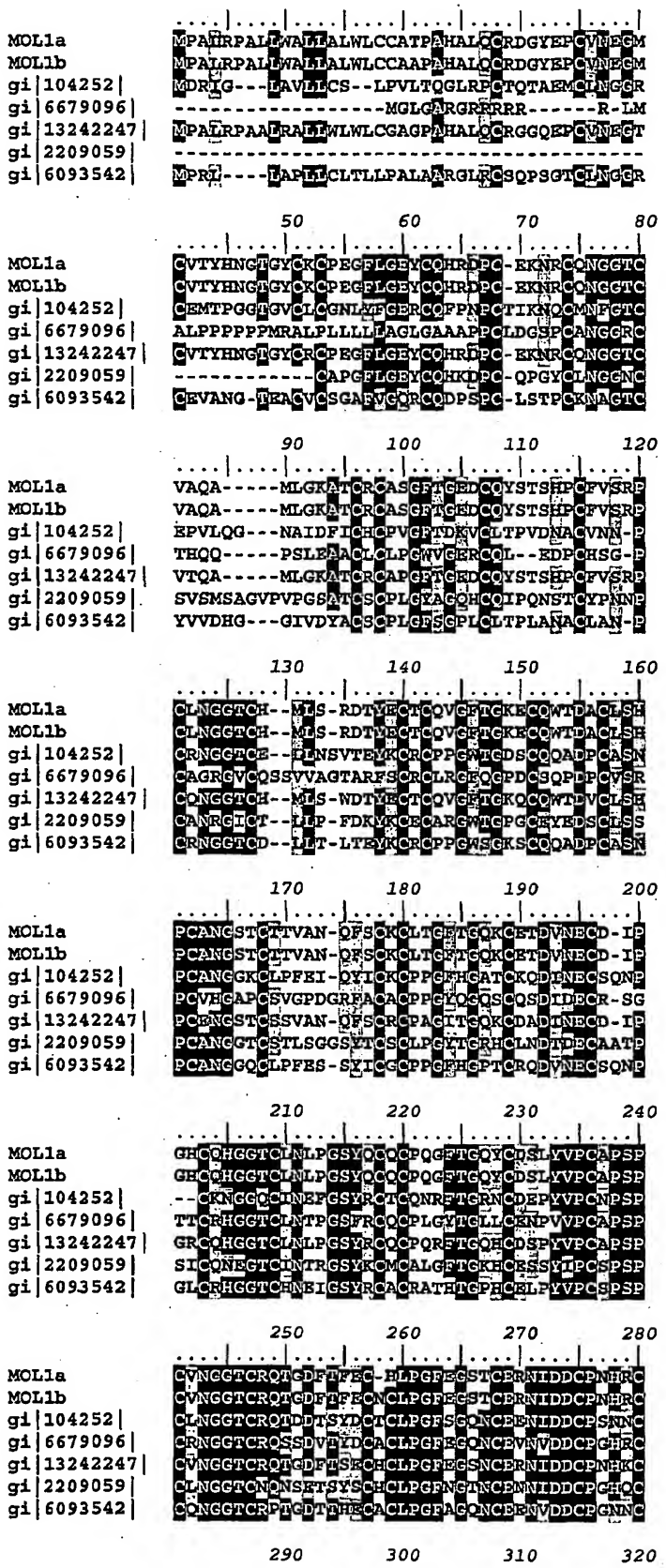
A ClustalW analysis comparing disclosed proteins of the invention with related OR protein sequences is given in Table 1H, with MOL1a shown on line 1 and MOL1b on line 2.

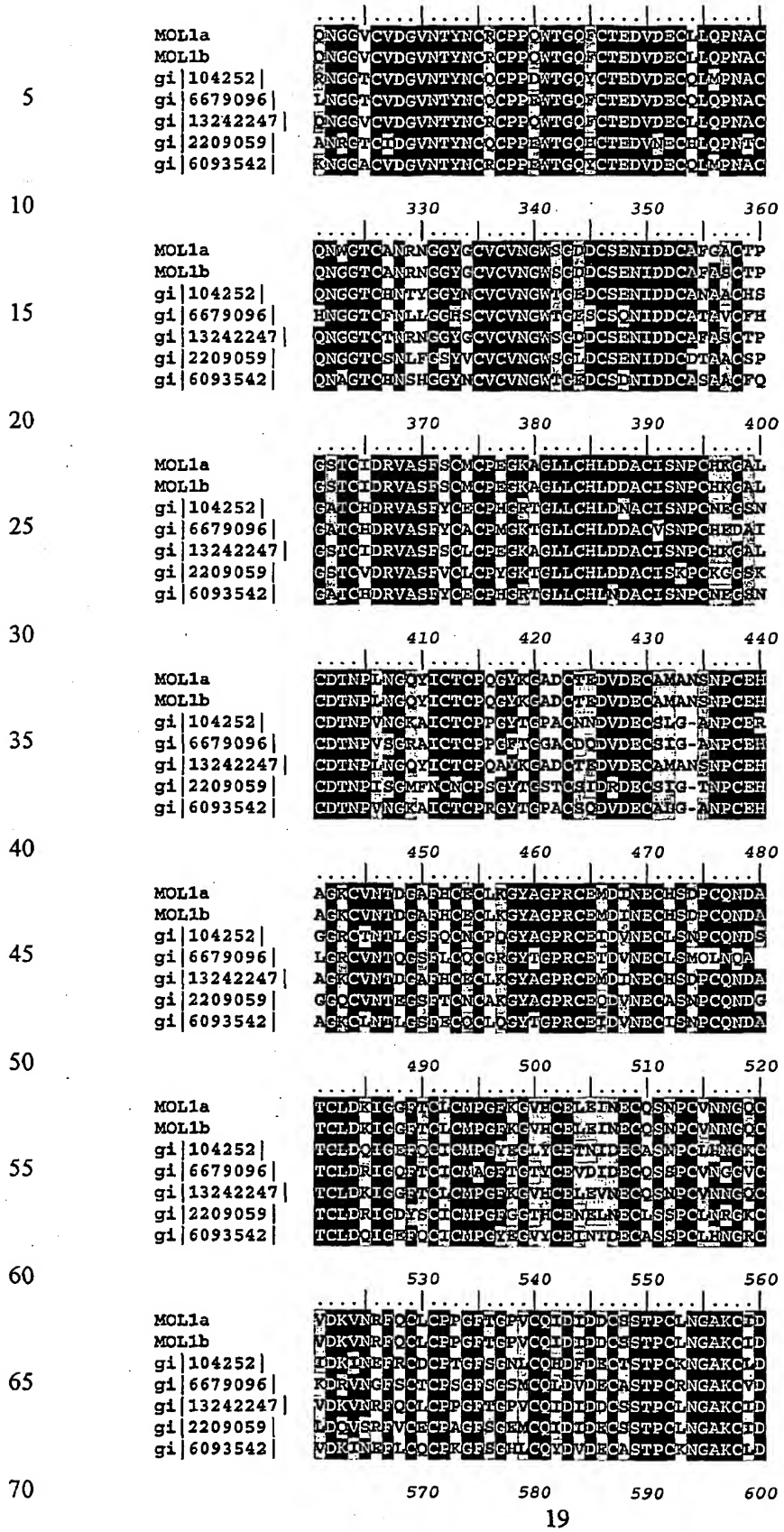
In the ClustalW alignment of the MOL1a and MOL1b proteins, as well as all other ClustalW analyses herein, the black outlined amino acid residues indicate regions of conserved sequence (*i.e.*, regions that may be required to preserve structural or functional properties), whereas non-highlighted amino acid residues are less conserved and can potentially be mutated to a much broader extent without altering protein structure or function. Residue differences between any MOLX variant sequences herein are written to show the residue in the "a" variant and the residue position with respect to the "a" variant. MOL residues in all following sequence alignments that differ between the individual MOL variants are highlighted with a box and marked with the (o) symbol above the variant residue in all alignments herein.

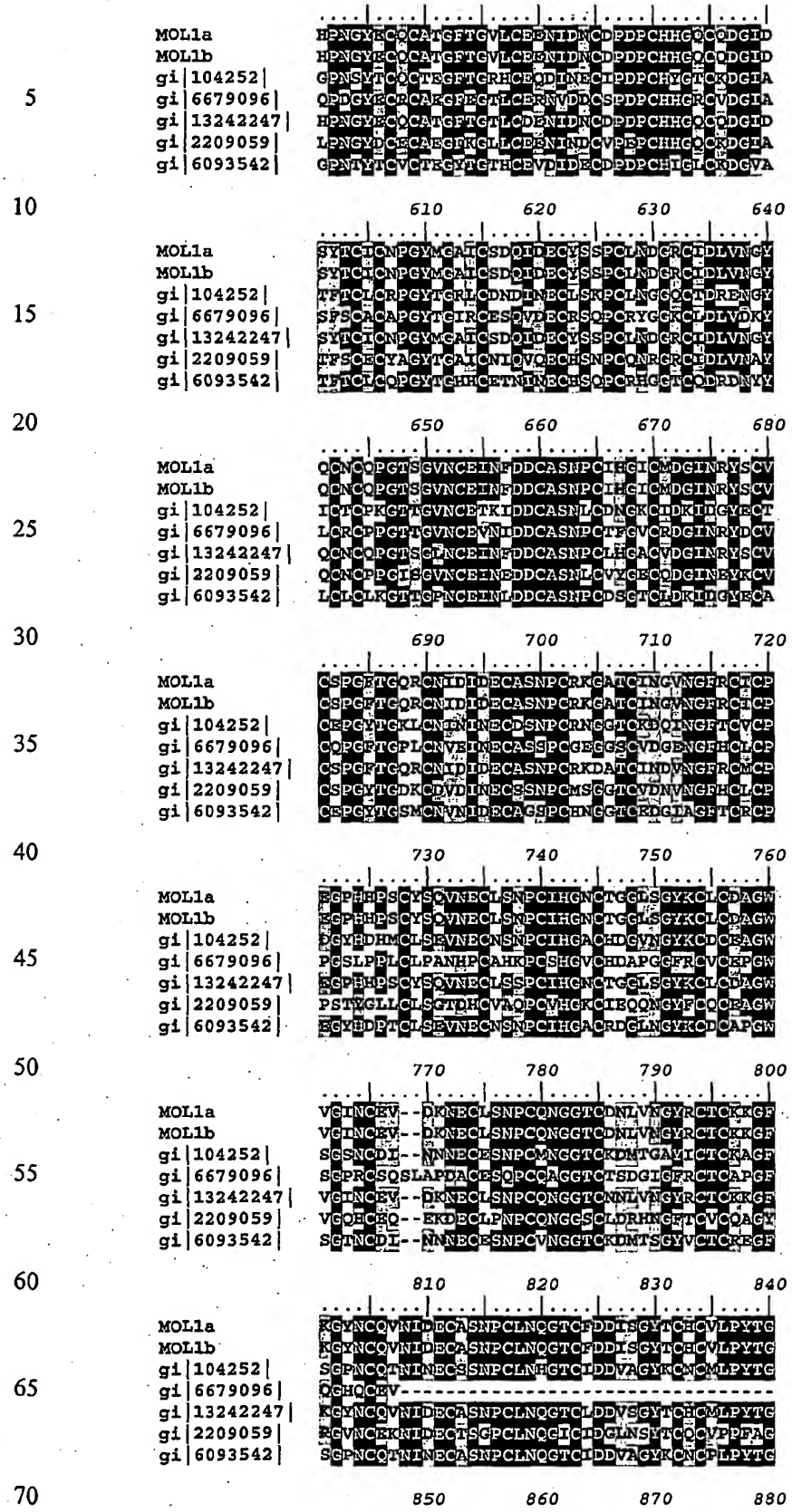
Table 1H. ClustalW Analysis of MOL1a

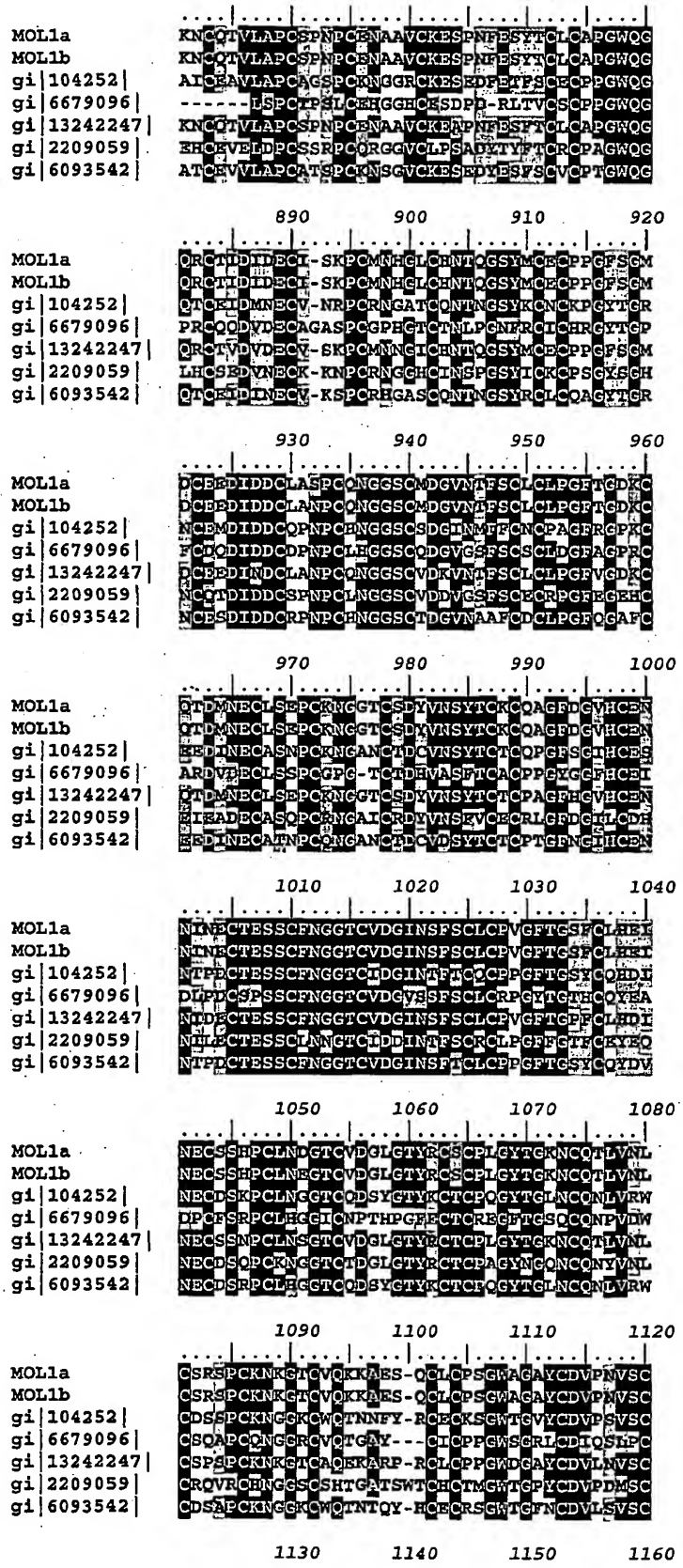
1) Novel MOL1a (SEQ ID NO:2)
2) Novel MOL1b (SEQ ID NO:4)
3) gi|104252|pir||A35844 Xotch protein - African clawed frog (SEQ ID NO:25)
4) gi|6679096|ref|NP_032742.1| Notch gene homolog 3, (Drosophila) [Mus musculus] (SEQ ID NO:26)
5) gi|13242247|ref|NP_077334.1| Notch gene homolog 2, (Drosophila) [Rattus norvegicus] (SEQ ID NO:27)
6) gi|2209059|dbj|BAA20535.1| Notch 2 [Takifugu rubripes] (SEQ ID NO:28)
7) gi|6093542|sp|Q07008|NTC1_RAT NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 1 PRECURSOR (SEQ ID NO:29)

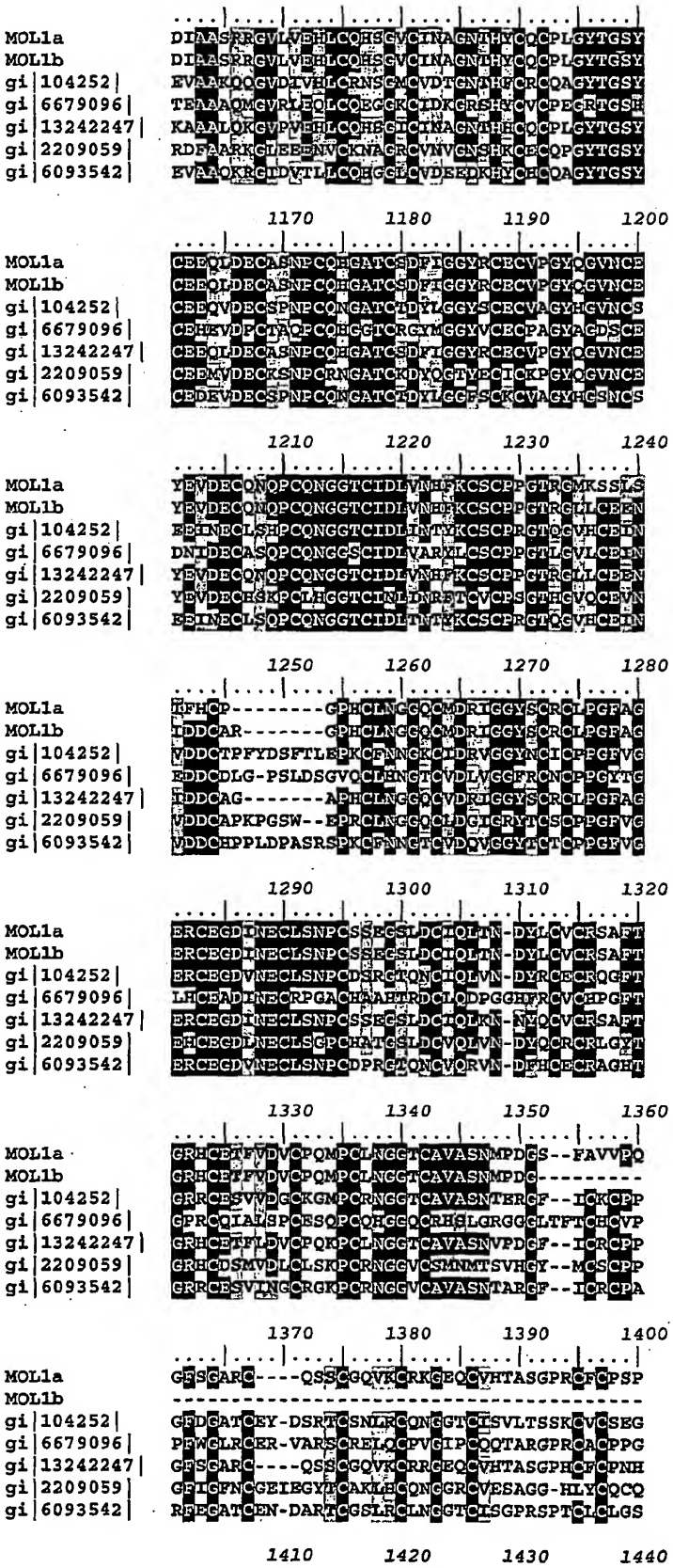
10 20 30 40

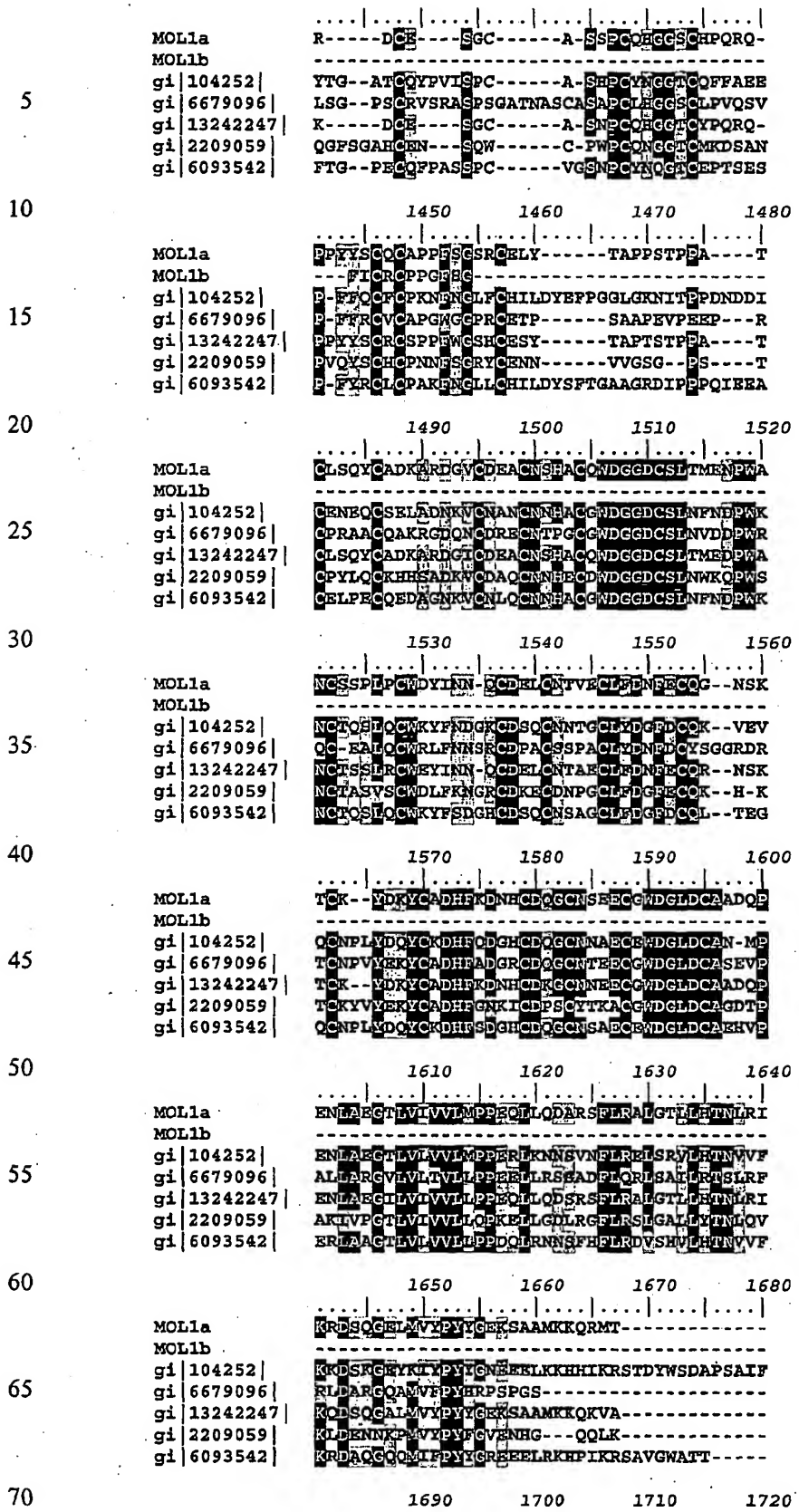

MOL1a
MOL1b
gi|104252|
gi|6679096|
gi|13242247|
gi|2209059|
gi|6093542|
50
60
70
80
90
100
110
120
130
140
150
160
170
180
190
200
210
220
230
240
250
260
270
280
290
300
310
320

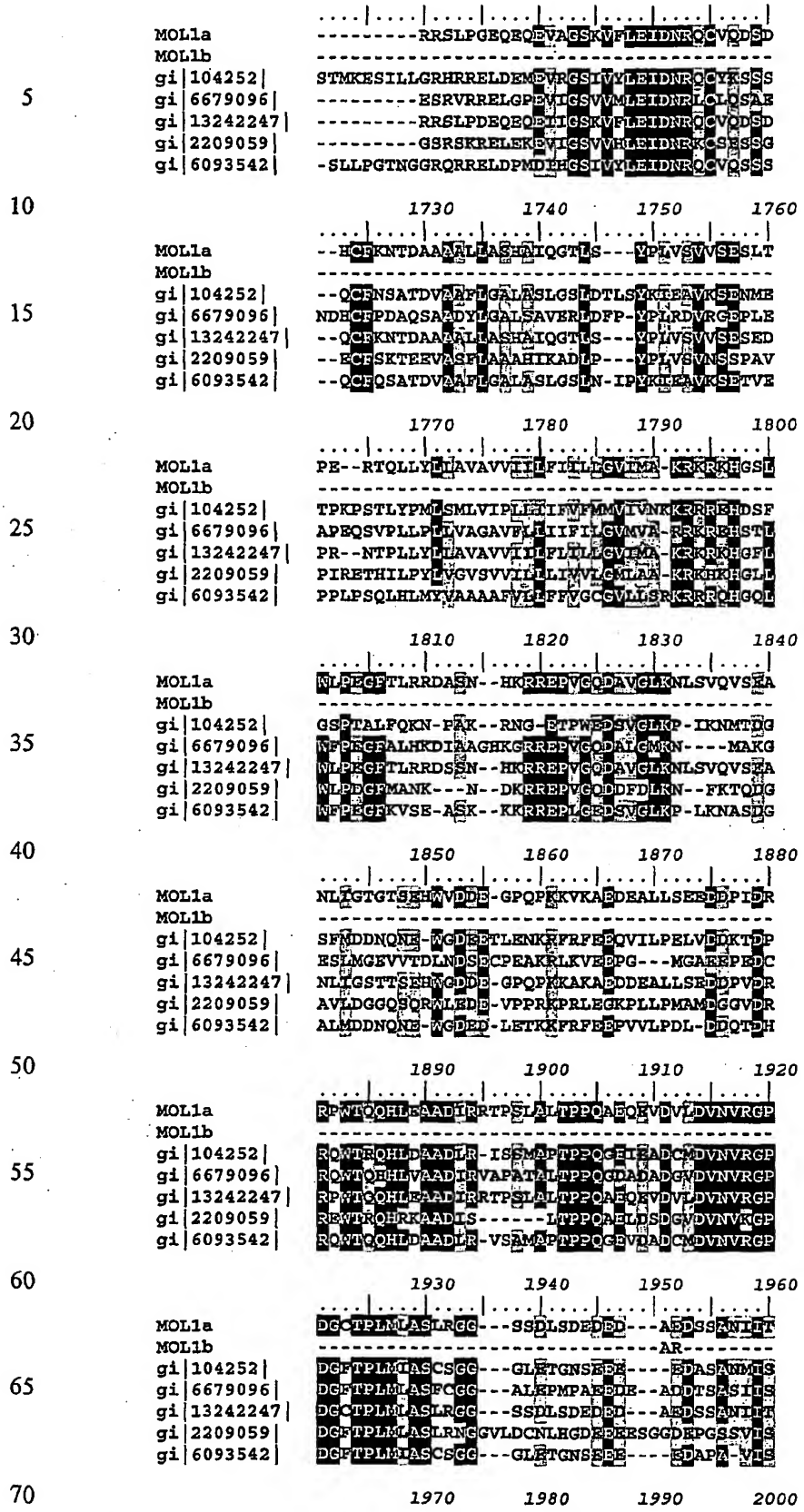

MOL1a
MOL1b
gi|104252|
gi|6679096|
gi|13242247|
gi|2209059|
gi|6093542|
330 340 350 360
370 380 390 400
410 420 430 440
450 460 470 480
490 500 510 520
530 540 550 560
570 580 590 600

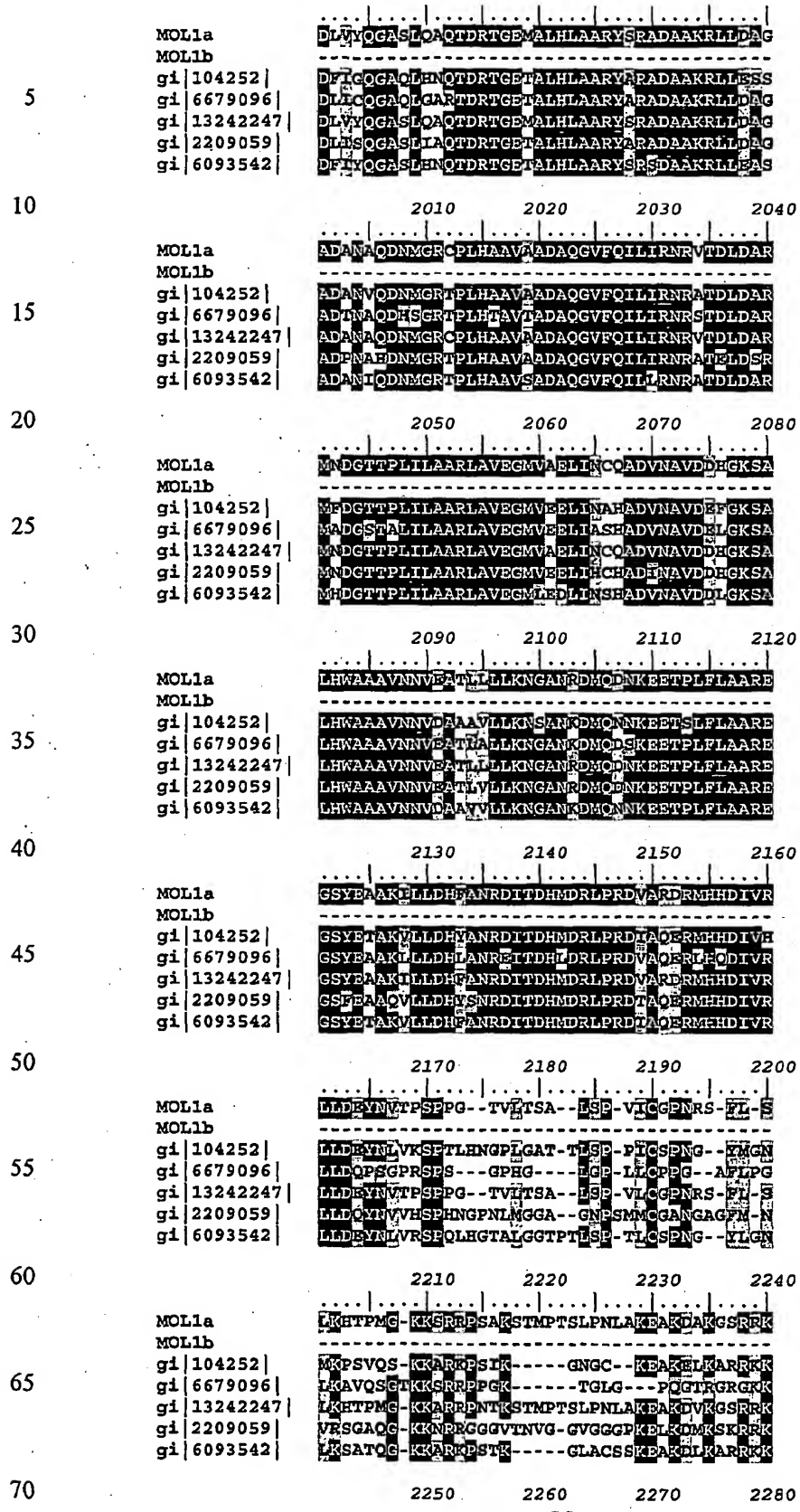

MOL1a
MOL1b
gi|104252|
gi|6679096|
gi|13242247|
gi|2209059|
gi|6093542|
610
620
630
640
650
660
670
680
690
700
710
720
730
740
750
760
770
780
790
800
810
820
830
840
850
860
870
880

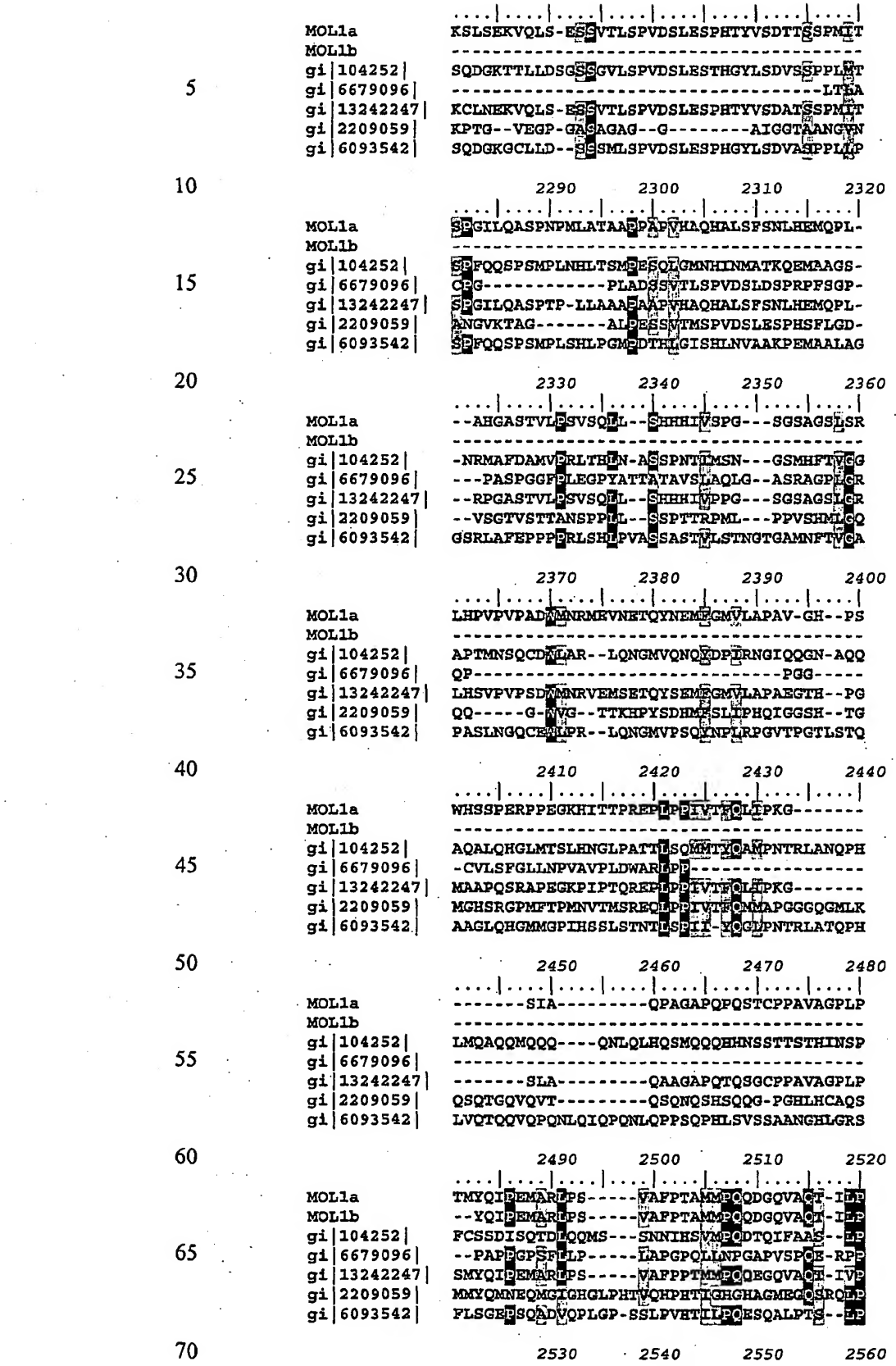

MOL1a
MOL1b
gi|104252|
gi|6679096|
gi|13242247|
gi|2209059|
gi|6093542|
890
900
910
920
930
940
950
960
970
980
990
1000
1010
1020
1030
1040
1050
1060
1070
1080
1090
1100
1110
1120
1130
1140
1150
1160

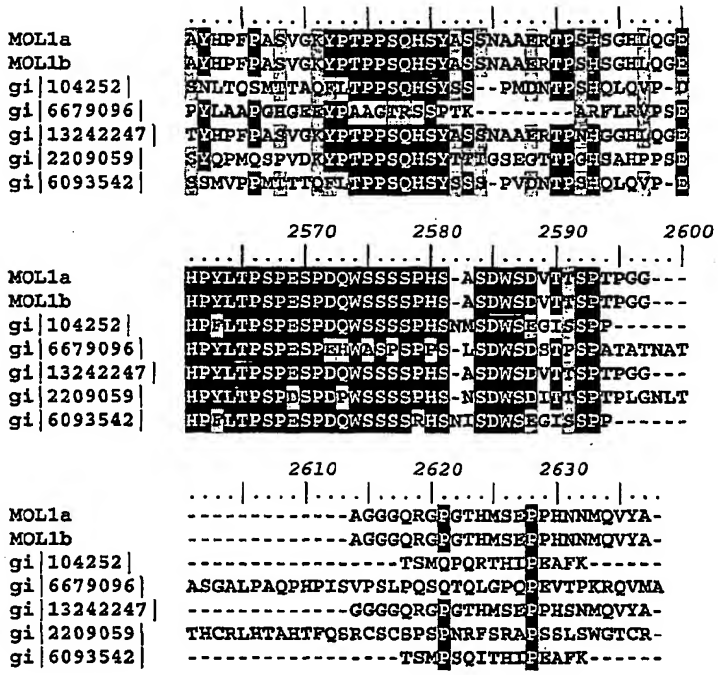

MOL1a
MOL1b
gi|104252|
gi|6679096|
gi|13242247|
gi|2209059|
gi|6093542|
1170
1180
1190
1200
1210
1220
1230
1240
1250
1260
1270
1280
1290
1300
1310
1320
1330
1340
1350
1360
1370
1380
1390
1400
1410
1420
1430
1440


MOL1a
MOL1b
gi|104252|
gi|6679096|
gi|13242247|
gi|2209059|
gi|6093542|
1450
1460
1470
1480
1490
1500
1510
1520
1530
1540
1550
1560
1570
1580
1590
1600
1610
1620
1630
1640
1650
1660
1670
1680
1690
1700
1710
1720


MOL1a
MOL1b
gi|104252|
gi|6679096|
gi|13242247|
gi|2209059|
gi|6093542|
1730 1740 1750 1760
1770 1780 1790 1800
1810 1820 1830 1840
1850 1860 1870 1880
1890 1900 1910 1920
1930 1940 1950 1960
1970 1980 1990 2000


MOL1a
MOL1b
gi|104252|
gi|6679096|
gi|13242247|
gi|2209059|
gi|6093542|
2010
2020
2030
2040
2050
2060
2070
2080
2090
2100
2110
2120
2130
2140
2150
2160
2170
2180
2190
2200
2210
2220
2230
2240
2250
2260
2270
2280


MOL1a
MOL1b
gi|104252|
gi|6679096|
gi|13242247|
gi|2209059|
gi|6093542|
2290 2300 2310 2320
2330 2340 2350 2360
2370 2380 2390 2400
2410 2420 2430 2440
2450 2460 2470 2480
2490 2500 2510 2520
2530 2540 2550 2560



When the sequences of the invention are referred to as MOL1, this refers to the sequences disclosed as MOL1a and MOL1b.

The presence of identifiable domains in MOL1, as well as all other MOLX proteins, was determined by searches using software algorithms such as PROSITE, DOMAIN, Blocks, Pfam, ProDomain, and Prints, and then determining the Interpro number by crossing the domain match (or numbers) using the Interpro website (http:www.ebi.ac.uk/ interpro). DOMAIN results, *e.g.*, for MOL1 as disclosed in Table 1I, were collected from the Conserved Domain Database (CDD) with Reverse Position Specific BLAST analyses. This BLAST analysis software samples domains found in the Smart and Pfam collections. For Table 1I and all successive DOMAIN sequence alignments, fully conserved single residues are indicated by black shading and "strong" semi-conserved residues are indicated by grey shading. The "strong" group of conserved amino acid residues may be any one of the following groups of amino acids: STA, NEQK, NHQK, NDEQ, QHRK, MILV, MILF, HY, FYW.

Tables 1I-1N list the domain description from DOMAIN analysis results against MOL1. The region from amino acid residue 1416 through 1454 (SEQ ID NO:2) most probably ($E = 1e^{-6}$) contains a domain found in Notch and Lin-12, aligned here in table 1I. Notch and Lin-12 are both involved in organismal development, cell cycle, and apoptosis. The region from amino acid residue 1504 through 1532 (SEQ ID NO:2) most probably ($E = 2e^{-4}$) also contains a domain found in Notch and Lin-12, aligned here in table 1J. The region from

amino acid residue 1875 through 1906 (SEQ ID NO:2) most probably ($E = 6e^{-5}$) contains an Ank repeat, aligned here in table 1K. Ank repeats are markers for the protein ankyrin which is involved in cell adhesion and contact inhibition. The region from amino acid residue 1974 through 2006 (SEQ ID NO:2) most probably ($E = 2e^{-4}$) also contains an Ank repeat, aligned here in table 1L. The region from amino acid residue 182 through 215 (SEQ ID NO:2) most probably ($E = 1e^{-4}$) contains a Calcium binding EGF-like domain, aligned here in table 1M. EGF is a growth factor that modulates the proliferation of many cell types. The region from amino acid residue 872 through 908 (SEQ ID NO:2) most probably ($E = 9e^{-4}$) also contains a Calcium binding EGF-like domain, aligned here in table 1N. This indicates that the MOL1 sequence has properties similar to those of other proteins known to contain these domains.

Table 1I. Domain Analysis of MOL1

gnl|Smart|smart00004, NL, Domain found in Notch and Lin-12; The Notch protein is essential for the proper differentiation of the Drosophila ectoderm. This protein contains 3 NL domain
CD-Length = 39 residues, 100.0% aligned
Score = 50.1 bits (118), Expect = 1e-06

```
                          10        20        30        40
                  ....|....|....|....|....|....|....|....|..
MOL1_1            PSTPPATCL---SQYCADKARDGVCDEACNSHACQNDGGDCS
Smart|smart00004  PQDPWSRCE---DAQCWDKFGDGVCDEECNNAECLWDGGDCS (SEQ ID NO:83)
```

Table 1J. Domain Analysis of MOL1

gnl|Smart|smart00004, NL, Domain found in Notch and Lin-12; The Notch protein is essential for the proper differentiation of the Drosophila ectoderm. This protein contains 3 NL domains
CD-Length = 39 residues, only 74.4% aligned
Score = 43.1 bits (100), Expect = 2e-04

```
                          10        20        30        40
                  ....|....|....|....|....|....|....|....|..
MOL1_2            PECQGNSKTCKYDKYCADHFKDNHCDQGCNSEECGWDGLDCA
Smart|smart00004  PQDPWSRCE---DAQCWDKFGDGVCDEECNNAECLWDGGDCS (SEQ ID NO:84)
```

Table 1K. Domain Analysis of MOL1

gnl|Pfam|pfam00023, ank, Ank repeat
CD-Length = 33 residues, 97.0% aligned
Score = 44.7 bits (104), Expect = 6e-05

```
                        10        20        30        40
                ....|....|....|....|....|....|....|....|....|
MOL1_3          TGEMALHLAARYS----------RADAARRLLD-AGADANAQDN
Pfam|pfam00023  DGNTPLHLAARNG----------HLEVVKLLLE-AGADVNARDK (SEQ ID NO:85)
```

Table 1L. Domain Analysis of MOL1

gnl|Pfam|pfam00023, ank, Ank repeat
CD-Length = 33 residues, 100.0% aligned
Score = 43.1 bits (100), Expect = 2e-04

```
                        10        20        30
                ....|....|....|....|....|....|....
```

Table 1M. Domain Analysis of MOL1

gnl|Smart|smart00179, EGF_CA, Calcium-binding EGF-like domain
CD-Length = 41 residues, 85.4% aligned
Score = 43.5 bits (101), Expect = 1e-04

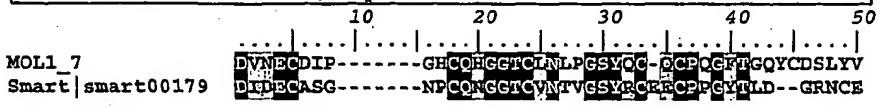
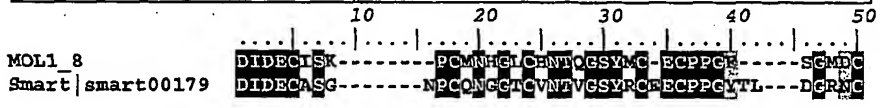
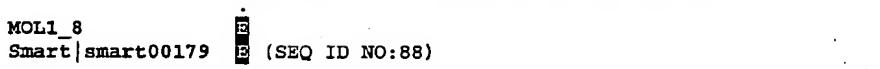


Table 1N. Domain Analysis of MOL1

gnl|Smart|smart00179, EGF_CA, Calcium-binding EGF-like domain
CD-Length = 41 residues, 100.0% aligned
Score = 40.8 bits (94), Expect = 9e-04





Uses of the Compositions of the Invention

The protein similarity information, expression pattern, cellular localization, and map location for the protein and nucleic acid disclosed herein suggest that MOL1 may have important structural and/or physiological functions characteristic of the EGF-like domain containing protein family. Therefore, the nucleic acids and proteins of the invention are useful in potential diagnostic and therapeutic applications and as a research tool. These include serving as a specific or selective nucleic acid or protein diagnostic and/or prognostic marker, wherein the presence or amount of the nucleic acid or the protein are to be assessed. These also include potential therapeutic applications such as the following: (i) a protein therapeutic, (ii) a small molecule drug target, (iii) an antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), (iv) a nucleic acid useful in gene therapy (gene delivery/gene

ablation), (v) an agent promoting tissue regeneration *in vitro* and *in vivo*, and (vi) a biological defense weapon.

The MOL1 nucleic acids and proteins have applications in the diagnosis and/or treatment of various diseases and disorders. For example, the compositions of MOL1 may have efficacy for the treatment of patients suffering from endometriosis, fertility disorders, cancer, hypercoagulation, idiopathic thrombocytopenic purpura, immunodeficiencies, systemic lupus erythematosus, asthma, emphysema, scleroderma, allergy, ARDS, Von Hippel-Lindau (VHL) syndrome, Alzheimer's disease, stroke, tuberous sclerosis, hypercalceimia, Parkinson's disease, Huntington's disease, cerebral palsy, epilepsy, multiple sclerosis, ulcers, ataxia-telangiectasia, leukodystrophies, behavioral disorders, addiction, anxiety, pain, neuroprotection, osteoporosis, hypercalceimia, arthritis, ankylosing spondylitis, scoliosis, diabetes, autoimmune disease, myasthenia gravis, muscular dystrophy, renal artery stenosis, interstitial nephritis, glomerulonephritis, polycystic kidney disease, systemic lupus erythematosus, renal tubular acidosis, IgA nephropathy, hypercalceimia, Lesch-Nyhan syndrome, developmental disorders, growth disorders, and/or wounds, as well as other diseases, disorders and conditions. The reactivation of the Notch signaling pathway during wound healing has been demonstrated and the similarity between developmental and regenerative processes has been suggested (*Exp Cell Res* 1999 Feb 1;246(2):312-8).

These materials are further useful in the generation of antibodies that bind immuno-specifically to the novel MOL1 substances for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-MOLX Antibodies" section below. The disclosed MOL1 protein has multiple hydrophilic regions, each of which can be used as an immunogen. In one embodiment, a contemplated MOL1 epitope is from about amino acids 10 to 150. In another embodiment, a MOL1 epitope is from about amino acids 160 to 190. In additional embodiments, MOL1 epitopes are from about amino acids 195 to 350, 400 to 525, 550 to 575, 590 to 600, 610 to 650, 780 to 880, 900 to 1000, 1100 to 1150, 1175 to 1200, 1225 to 1300, and from about amino acids 1380 to 1500. These novel proteins can also be used to develop assay systems for functional analysis.

MOL2

MOL2 is a novel insulin-like growth factor binding protein-like protein. The novel nucleic acid of 2631 nucleotides, (SC98428706_EXT, SEQ ID NO:5) encoding a novel insulin-like growth factor binding protein-like protein is shown in Table 2A. The start and stop codons are in bold.

Table 2A. MOL2 Nucleotide Sequence (SEQ ID NO:5)

ATGATTTACATACAAGTAATTTTTCAAGTAATGACCATTGAAAAAATGTTTTCTTTTTATTTTTAGATTATTTCTC
TTTATTCAGAAGCATACAGTTGTTTGCTGATTGCAAGAAGATGTTTCTGTGGCTGTTTCTGATTTTGTCAGCCCTGA
TTTCTTCGACAAATGCAGATTCTGACATATCGGTGGAAATTTGCAATGTGTGTTCCTGCGTGTCAGTTGAGAATGTG
CTCTATGTCAACTGTGAGAAGGTTTCAGTCTACAGACCAAATCAGCTGAAACCACCTTGGTCTAATTTTTATCACCT
CAATTTCCAAAATAATTTTTTAAATATTCTGTATCCAAATACATTCTTGAATTTTTCACATGCAGTCTCCCTGCATC
TGGGGAATAATAAACTGCAGAACATTGAGGGAGGAGCCTTTCTTGGGCTCAGTGCATTAAAGCAGTTGCACTTGAAC
AACAATGAATTAAAGATTCTCCGAGCTGACACTTTCCTTGGCATAGAGAACTTGGAGTATCTCCAGGCTGACTACAA
TTTAATCAAGTATATTGAACGAGGAGCCTTCAATAAGCTCCACAAACTGAAAGTTCTCATTCTTAATGACAATCTGA
TTTCATTCCTTCCTGATAATATTTTCCGATTCGCATCTTTGACCCATCTGGATATACGAGGGAACAGAATCCAGAAG
CTCCCTTATATCGGGGGTTCTGGAACACATTGGCCGTGTCGTTGAATTGCAACTGGAAGATAACCCTTGGAACTGTAG
CTGTGATTTATTGCCCTTAAAAGCTTGGCTGGAGAACATGCCATATAACATTTACATAGGAGAAGCTATCTGTGAAA
CTCCCAGTGACTTATATGGAAGGCTTTTAAAAGAAACCAACAAACAAGAGCTATGTCCCATGGGCACCGGCAGTGAT
TTTGACGTGCGCATCCTGCCTCCATCTCAGCTGGAAAATGGCTACACCACTCCCAATGGTCACACTACCCAAACATC
TTTACACAGATTAGTAACTAAACCACCAAAAACAACAAATCCTTCCAAGATCTCTGGAATCGTTGCAGGCAAAGCCC
TCTCCAACCGCAATCTCAGTCAGATTGTGTCTTACCAAACAAGGGTGCCTCCTCTAACACCTTGCCCGGCACCTTGC
TTCTGCAAAACACACCCTTCAGATTTGGGACTAAGTGTGAACTGCCAAGAGAAAAATATACAGTCTATGTCTGAACT
GATACCGAAACCTTTAAATGCGAAGAAGCTGCACGTCAATGGCAATAGCATCAAGGATGTGGACGTATCAGACTTCA
CTGACTTTGAAGGACTGGATTTGCTTCATCTAGGCAGCAATCAAATTACAGTGATTAAGGGAGACGTATTTCACAAT
CTCACTAATTTACGCAGGCTATATCTCAATGGCAATCAAATTGAGAGACTCTATCCTGAAATATTTTCAGGTCTTCA
TAACCTGCAGTATCTGTATTTGGAATACAATTTGATTAAGGAAATCTCAGCAGGCACCTTTGACTCCATGCCAAATT
TGCAGTTACTGTACTTAAACAATAATCTCCTAAAGAGCCTGCCTGTTTACATCTTTTCCGGAGCACCCTTAGCTAGA
CTGAACCTGAGGAACAACAAATTCATGTACCTGCCTGTCAGTGGGGTCCTTGATCAGTTGCAATCTCTTACACAGAT
TGACTTGGAGGGCAACCCATGGGACTGTACTTGTGACTTGGTGGCATTAAAGCTGTGGGTGGAGAAGTTGAGCGACG
GGATTGTTGTGAAAGAACTGAAATGTGAGACGCCTGTTCAGTTTGCCAACATTGAACTGAAGTCCCTCAAAAATGAA
ATCTTATGTCCCAAACTTTTAAATAAGCCGTCTGCACCATTCACAAGCCCTGCACCTGCCATTACATTCACCACTCC
TTTGGGTCCCATTCGAAGTCCTCCTGGTGGGCCAGTGCCTCTGTCTATTTTAATCTTAAGTATCTTAGTGGTCCTCA
TTTTAACGGTGTTTGTTGCTTTTTGCCTTCTTGTTTTTGTCCTGCGACGCAACAAGAAACCCACAGTGAAGCACGAA
GGCCTGGGGAATCCTGACTGTGGCTCCATGCAGCTGCAGCTAAGGAAGCATGACCACAAAACCAATAAAAAAGATGG
ACTGAGCACAGAAGCTTTCATTCCACAAACTATAGAACAGATGAGCAAGAGCCACACTTGTGGCTTGAAAGAGTCAG
AAACTGGGTTCATGTTTTCAGATCCTCCAGGACAGAAAGTTGTTATGAGAAATGTGGCCGACAAGGAGAAAGATTTA
TTACATGTAGATACCAGGAAGAGACTGAGCACAATTGATGAGCTGGATGAATTATTCCCTAGCAGGGATTCCAATGT
GTTTATTCAGAATTTTCTTGAAAGCAAAAAGGAGTATAATAGCATAGGTGTCAGTGGCTTTGAGATCCGCTATCCAG
AAAAACAACCAGACAAAAAAAGTAAGAAGTCACTGATAGGTGGCAACCACAGTAAAATTGTTGTGGAACAAAGGAAG
AGTGAGTATTTTGAACTGAAGGCGAAACTGCAGAGTTCCCCTGACTACCTACAGGTCCTTGAGGAGCAAACAGCTTT
GAACAAGATCTAG

An open reading frame (ORF) for MOL2 was identified from nucleotides 1 to 2628. The disclosed MOL2 polypeptide (SEQ ID NO:6) encoded by SEQ ID NO:5 has 876 amino acid residues and is presented using the one-letter code in Table 2B. The SignalP, Psort and Hydropathy profile of MOL2 indicate that this sequence does have a signal peptide localized between amino acids 57 and 58 (TNA-DS) and is likely to be localized to the plasma membrane (0.4600 certainty). Therefore it is likely that MOL2 is available at the appropriate sub-cellular localization and hence accessible for the therapeutic uses described in this application.

Table 2B. Encoded MOL2 protein sequence (SEQ ID NO:6).

MIYIQVIFQVMTIEKMFSFYFLDYFSLFRSIQLFADCKKMFLWLFLILSALISSTNADSDISVEICNVCSCVSVENV
LYVNCEKVSVYRPNQLKPPWSNFYHLNFQNNFLNILYPNTFLNFSHAVSLHLGNNKLQNIEGGAFLGLSALKQLHLN
NNELKILRADTFLGIENLEYLQADYNLIKYIERGAFNKLHKLKVLILNDNLISFLPDNIFRFASLTHLDIRGNRIQK
LPYIGVLEHIGRVVELQLEDNPWNCSCDLLPLKAWLENMPYNIYIGEAICETPSDLYGRLLKETNKQELCPMGTGSD
FDVRILPPSQLENGYTTPNGHTTQTSLHRLVTKPPKTTNPSKISGIVAGKALSNRNLSQIVSYQTRVPPLTPCPAPC
PCKTHPSDLGLSVNCQEKNIQSMSELIPKPLNAKKLHVNGNSIKDVDVSDFTDFEGLDLLHLGSNQITVIKGDVFHN
LTNLRRLYLNGNQIERLYPEIFSGLHNLQYLYLEYNLIKEISAGTFDSMPNLQLLYLNNNLLKSLPVYIFSGAPLAR
LNLRNNKFMYLPVSGVLDQLQSLTQIDLEGNPWDCTCDLVALKLWVEKLSDGIVVKELKCETPVQFANIELKSLKNE
ILCPKLLNKPSAPFTSPAPAITFTTPLGPIRSPPGGPVPLSILILSILVVLILTVFVAFCLLVFVLRRNKKPTVKHE
GLGNPDCGSMQLQLRKHDHKTNKKDGLSTEAFIPQTIEQMSKSHTCGLKESETGFMFSDPPGQKVVMRNVADKEKDL
LHVDTRKRLSTIDELDELFPSRDSNVFIQNPLESKKEYNSIGVSGFEIRYPEKQPDKKSKKSLIGGNHSKIVVEQRK
SEYFELKAKLQSSPDYLQVLEEQTALNKI

The MOL2 nucleic acid sequence, localized on the q26.3-28 region of the X chromosome, has 532 of 854 bases (62%) identical to a *Homo sapiens* Insulin-like growth factor binding protein-like protein mRNA (GENBANK-ID:AB020655|acc:AB020655).

The full amino acid sequence of the protein of the invention was found to have 318 of 672 amino acid residues (47%) identical to, and 445 of 672 residues (66%) similar to, the 977 amino acid residue Insulin-like growth factor binding protein-like protein from *Homo sapiens* (SPTREMBL-ACC:O94933).

MOL2 expression in different tissues was examined through TaqMan as described below in Example 1.

Other BLAST results including the sequences used for ClustalW analysis are presented in Table 2C.

Table 2C. BLAST results for MOL2

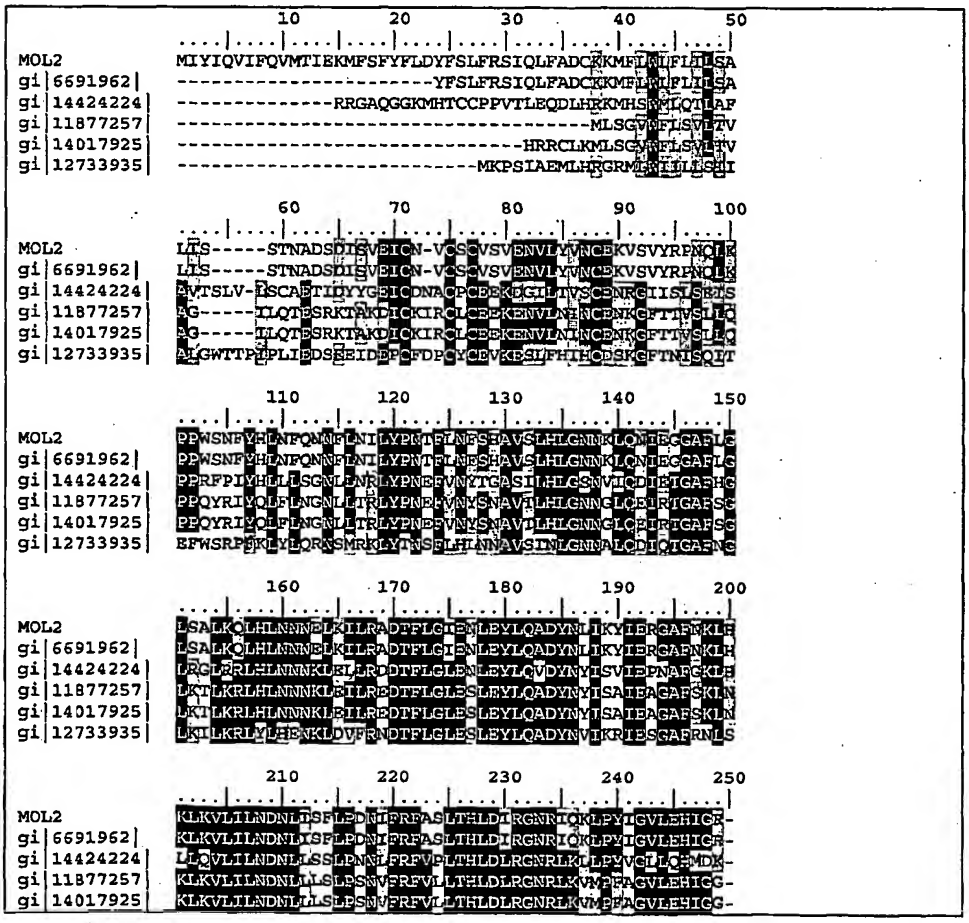
Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi|6691962|emb|CAB65788.1| (AL080239)	bG256O22.1 (similar to IGFALS (insulin-like growth factor binding protein, acid labile subunit)) [Homo sapiens]	853	853/853 (100%)	853/853 (100%)	0.0
gi|14424224|sp|O94991|Y918_HUMAN	HYPOTHETICAL PROTEIN KIAA0918 [Homo Sapiens]	966	347/699 (49%)	470/699 (66%)	1 e-173
gi|11877257|emb|CAC18888.1| (AL109653)	bG115M3.1 (novel protein) [Homo sapiens]	845	372/868 (42%)	516/868 (58%)	1 e-172

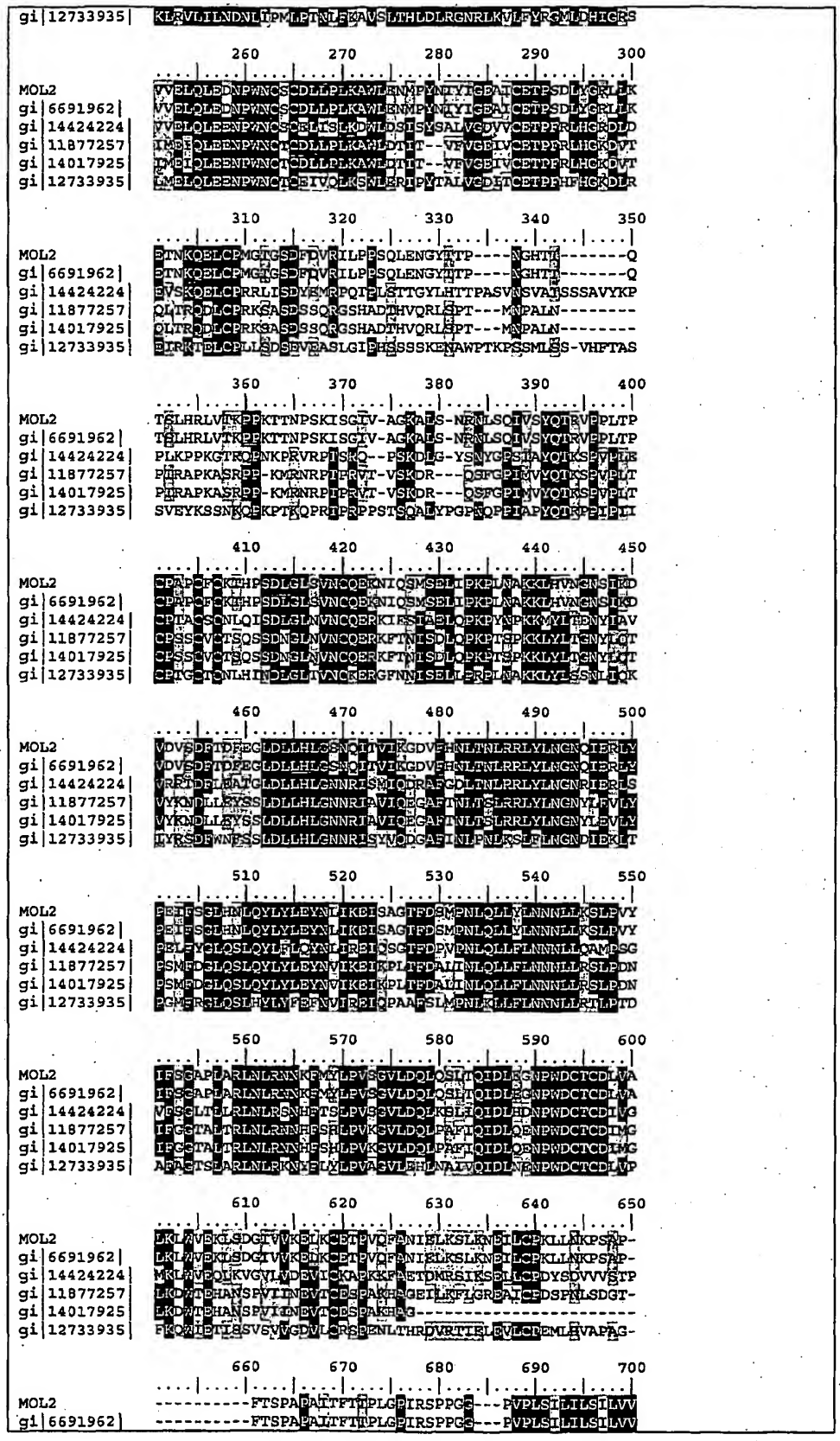
gi|12733935|ref|XP_011654.1|	KIAA0848 protein [Homo sapiens]	977	313/675 (46%)	438/675 (64%)	e-150
gi|7662336|ref|NP_055741.1|	KIAA0848 protein [Homo sapiens]	977	313/675 (46%)	438/675 (64%)	e-150

This information is presented graphically in the multiple sequence alignment given in Table 2D (with MOL2 being shown on line 1) as a ClustalW analysis comparing MOL2 with related protein sequences.

Table 2D. Information for the ClustalW proteins:

1) Novel MOL2 (SEQ ID NO:6)
2) gi|6691962|emb|CAB65788.1| (AL080239) insulin-like growth factor binding protein, acid labile subunit)) [Homo sapiens] (SEQ ID NO:30)
3) gi|14424224|sp|O94991|Y918_HUMAN HYPOTHETICAL PROTEIN KIAA0918 [Homo Sapiens] (SEQ ID NO:31)
4) gi|11877257|emb|CAC18888.1| (AL109653) bG115M3.1 (novel protein) [Homo sapiens] (SEQ ID NO:32)
5) gi|12733935|ref|XP_011654.1| KIAA0848 protein [Homo sapiens] (SEQ ID NO:33)
6) gi|7662336|ref|NP_055741.1| KIAA0848 protein [Homo sapiens] (SEQ ID NO:34)




gi|12733935|
260 270 280 290 300
MOL2
gi|6691962|
gi|14424224|
gi|11877257|
gi|14017925|
gi|12733935|
310 320 330 340 350
MOL2
gi|6691962|
gi|14424224|
gi|11877257|
gi|14017925|
gi|12733935|
360 370 380 390 400
MOL2
gi|6691962|
gi|14424224|
gi|11877257|
gi|14017925|
gi|12733935|
410 420 430 440 450
MOL2
gi|6691962|
gi|14424224|
gi|11877257|
gi|14017925|
gi|12733935|
460 470 480 490 500
MOL2
gi|6691962|
gi|14424224|
gi|11877257|
gi|14017925|
gi|12733935|
510 520 530 540 550
MOL2
gi|6691962|
gi|14424224|
gi|11877257|
gi|14017925|
gi|12733935|
560 570 580 590 600
MOL2
gi|6691962|
gi|14424224|
gi|11877257|
gi|14017925|
gi|12733935|
610 620 630 640 650
MOL2
gi|6691962|
gi|14424224|
gi|11877257|
gi|14017925|
gi|12733935|
660 670 680 690 700
MOL2
gi|6691962|

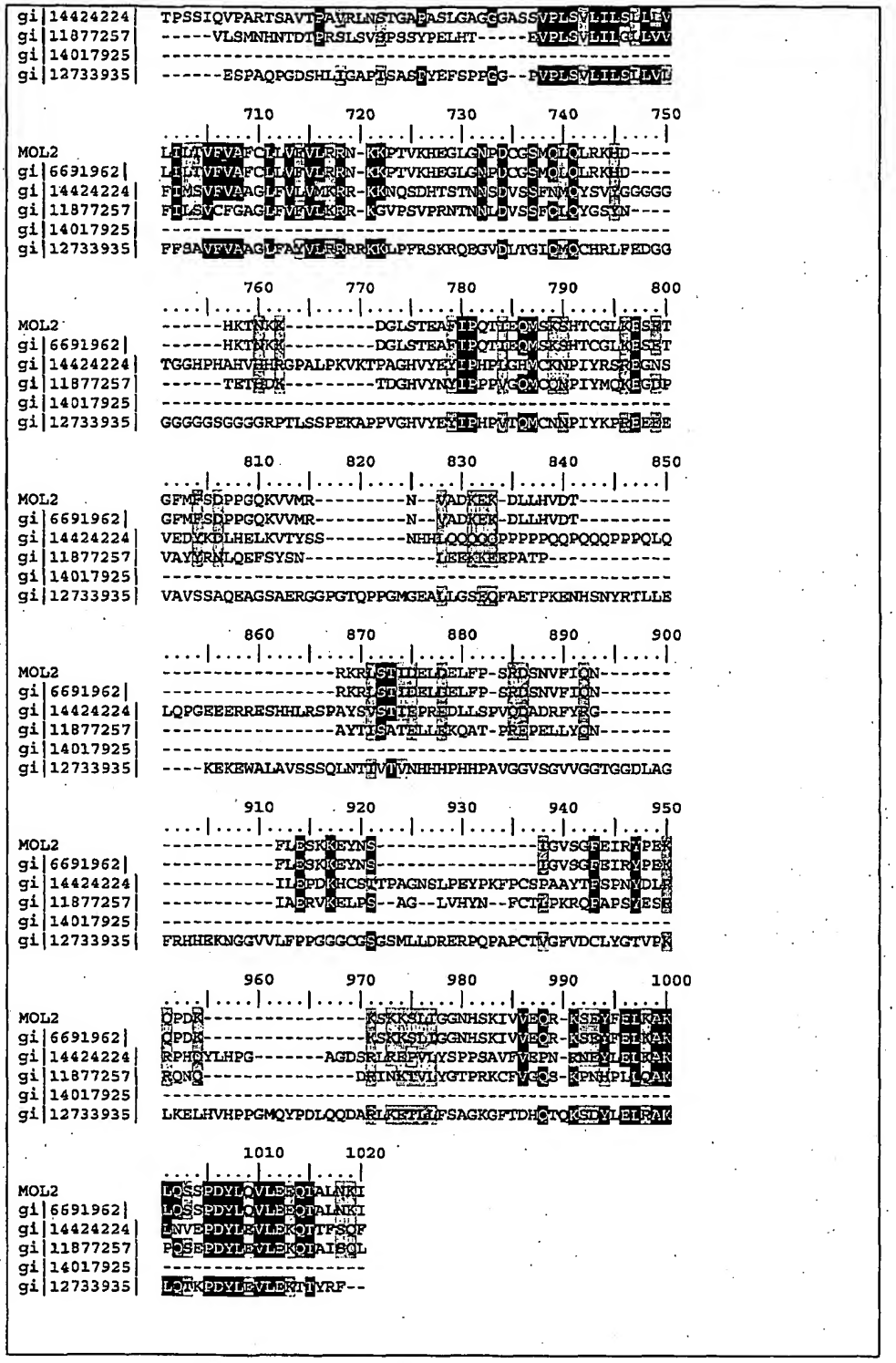
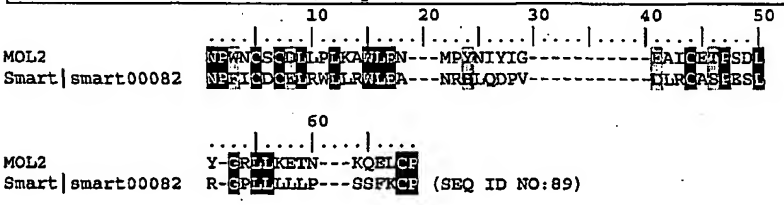


Table 2E lists the domain description from DOMAIN analysis results against MOL2. The region from amino acid residue 252 through 302 (SEQ ID NO:6) most probably ($E = 1e\text{-}^{6}$) contains a "leucine rich repeat C-terminal" domain, aligned here in Table 2E. This indicates

that the MOL2 sequence has properties similar to those of other proteins known to contain this domain.

Table 2E. Domain Analysis of MOL2

gnl|Smart|smart00082 LRRCT, Leucine rich repeat C-terminal domain
CD-Length = 51 residues, 100.0% aligned
Score = 48.9 bits (115), Expect = 1e-06



Chromosomal information

The Insulin-like growth factor binding protein-like protein disclosed in this invention maps to chromosome Xq26.3-28.

Tissue expression

MOL2 is expressed in at least the following tissues: adrenal gland, lymphatic tissues, and heart. Other tissues known to express insulin-like growth factor binding proteins are likely.

Uses of the Compositions of the Invention

The expression pattern, map location and protein similarity information for MOL2 suggest that this a Insulin-like growth factor binding protein-like protein may function as a member of the Insulin-like growth factor binding protein-like protein family. Therefore, the MOL2 nucleic acids and proteins are useful in potential therapeutic applications implicated, for example but not limited to, in various pathologies /disorders as described below and/or other pathologies/disorders. Potential therapeutic uses for MOL2 are, for example but not limited to, the following: (i) Protein therapeutic, (ii) small molecule drug target, (iii) antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), (iv) diagnostic and/or prognostic marker, (v) gene therapy (gene delivery/gene ablation), (vi) research tools, and (vii) tissue regeneration *in vitro* and *in vivo* (regeneration for all these tissues and cell types composing these tissues and cell types derived from these tissues).

The MOL2 nucleic acids and proteins are useful in potential therapeutic applications implicated in various diseases and disorders described below and/or other pathologies and disorders. For example, but not limited to, a cDNA encoding the a Insulin-like growth factor binding protein-like protein may be useful in gene therapy, and the a Insulin-like growth factor

binding protein-like protein may be useful when administered to a subject in need thereof. By way of nonlimiting example, the compositions of the present invention will have efficacy for treatment of patients suffering from cancer, diabetes, cardiomyopathy, atherosclerosis, hypertension, congenital heart defects, aortic stenosis, atrial septal defect (ASD), atrioventricular (A-V) canal defect, ductus arteriosus, pulmonary stenosis, subaortic stenosis, ventricular septal defect (VSD), valve diseases, tuberous sclerosis, scleroderma, obesity, transplantation, hemophilia, hypercoagulation, idiopathic thrombocytopenic purpura, autoimmume disease, allergies, immunodeficiencies, graft versus host disease (GVHD), lymphaedema, adrenoleukodystrophy, and/or congenital adrenal hyperplasia. MOL2, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

These materials are further useful in the generation of antibodies that bind immunospecifically to the novel MOL2 substances for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-MOLX Antibodies" section below. The disclosed MOL2 protein has multiple hydrophilic regions, each of which can be used as an immunogen. In one embodiment, a contemplated MOL2 epitope is from about amino acids 75 to 120. In another embodiment, a MOL2 epitope is from about amino acids 180 to 200. In additional embodiments, MOL2 epitopes are from about amino acids 280 to 380, 400 to 450, 475 to 500, and from about amino acids 680 to 850. These novel proteins can also be used to develop assay systems for functional analysis.

MOL3

MOL3a

An additional protein of the invention, referred to herein as MOL3a, is a human Semaphorin B-like protein. The novel nucleic acid of 2271 nucleotides (SC85516573_EXT, SEQ ID NO:7) encoding a novel olfactory receptor-like protein is shown in Table 3A. An open reading frame (ORF) was identified beginning with an ATG initiation codon at nucleotides 1-3 and ending with a TAA codon at nucleotides 2269-2271. The nucleotide sequence is presented in Table 3A with the start and stop codons are in bold letters.

Table 3A. MOL3a Nucleotide Sequence (SEQ ID NO:7)

ATGGCCCTCCCAGCCCTGGGCCTGGACCCCTGGAGCCTCCTGGGCCTTTTCCTCTTCCAACTGCTTC
AGCTGCTGCTGCCGACGACGACCGCGGGGGGAGGCGGGCAGGGGCCCATGCCCAGGGTCAGATACTA

TGCAGGGGATGAACGTAGGGCACTTAGCTTCTTCCACCAGAAGGGCCTCCAGGATTTTGACACTCTG
CTCCTGAGTGGTGATGGAAATACTCTCTACGTGGGGGCTCGAGAAGCCATTCTGGCCTTGGATATCC
AGGATCCAGGGGTCCCCAGGCTAAAGAACATGATACCGTGGCCAGCCAGTGACAGAAAAAAGAGTGA
ATGTGCCTTTAAGAAGAAGAGCAATGAGACACAGTGTTTCAACTTCATCCGTGTCCTGGTTTCTTAC
AATGTCACCCATCTCTACACCTGCGGCACCTTCGCCTTCAGCCCTGCTTGTACCTTCATTGAACTTC
AAGATTCCTACCTGTTGCCCATCTCGGAGGACAAGGTCATGGAGGGAAAAGGCCAAAGCCCCTTTGA
CCCCGCTCACAAGCATACGGCTGTCTTGGTGGATGGGATGCTCTATTCTGGTACTATGAACAACTTC
CTGGGCAGTGAGCCCATCCTGATGCGCACACTGGGATCCCAGCCTGTCCTCAAGACCGACAACTTCC
TCCGCTGGCTGCATCATGACGCCTCCTTTGTGGCAGCCATCCCTTCGACCCAGGTCGTCTACTTCTT
CTTCGAGGAGACAGCCAGCGAGTTTGACTTCTTTGAGAGGCTCCACACATCGCGGGTGGCTAGAGTC
TGCAAGAATGACGTGGGCGGCGAAAAGCTGCTGCAGAAGAAGTGGACCACCTTCCTGAAGGCCCAGC
TGCTCTGCACCCAGCCGGGGCAGCTGCCCTTCAACGTCATCCGCCACGCGGTCCTGCTCCCCGCCGA
TTCTCCCACAGCTCCCCACATCTACGCAGTCTTCACCTCCCAGTGGCAGGTTGGCGGGACCAGGAGC
TCTGCGGTTTGTGCCTTCTCTCTCTTGGACATTGAACGTGTCTTTAAGGGGAAATACAAAGAGTTGA
ACAAAGAAACTTCACGCTGGACTACTTATAGGGGCCCTGAGACCAACCCCCGGCCAGGCAGTTGCTC
AGTGGGCCCCTCCTCTGATAAGGCCCTGACCTTCATGAAGGACCATTTCCTGATGGATGAGCAAGTG
GTGGGGACGCCCCTGCTGGTGAAATCTGGCGTGGAGTATACACGGCTTGCAGTGGAGACAGCCCAGG
GCCTTGATGGGCACAGCCATCTTGTCATGTACCTGGGAACCAGTACAGGGTCGCTCCACAAGGCTGT
GGTAAGTGGGGACAGCAGTGCTCATCTGGTGGAAGAGATTCAGCTGTTCCCTGACCCTGAACCTGTT
CGCAACCTGCAGCTGGCCCCCACCCAGGGTGCAGTGTTTGTAGGCTTCTCAGGAGGTGTCTGGAGGG
TGCCCCGAGCCAACTGTAGTGTCTATGAGAGCTGTGTGGACTGTGTCCTTGCCCGGGACCCCCACTG
TGCCTGGGACCCTGAGTCCCGACTCTGCTCTCTTAGGAACTCCTGGAAGCAGGACATGGAGCGGGGG
AACCCAGAGTGGGCATGTGCCAGTGGCCCCATGAGCAGGAGCCTTCGGCCTCAGAGCCGCCCGCAAA
TCGTTAAAGAAGTCCTGGCTGTCCCCAACTCCATCCTGGAGCTCCCCTGCCCCCACCTGTCAGCCTT
GGCCTCTTATTATTGGAGTCATGGCCCAGCAGCAGTCCCAGAAGCCTCTTCCACTGTCTACAATGGC
TCCCTCTTGCTGATAGTGCAGGATGGAGTTGGGGGTCTCTACCAGTGCTGGGCAACTGAGAATGGCT
TTTCATACCCTGTGATCTCCTACTGGGTGGACAGCCAGGACCAGACCCTGGCCCTGGATCCTGAACT
GGCAGGCATCCCCCGGGAGCATGTGAAGGTCCCGTTGACCAGGGTCAGTGGTGGGGCCGCCCTGGCT
GCCCAGCAGTCCTACTGGCCCCACTTTGTCACTGTCACTGTCCTCTTTGCCTTAGTGCTTTCAGGAG
CCCTCATCATCCTCGTGGCCTCCCCATTGAGAGCACTCCGGGCTCGGGGCAAGGTTCAGGGCTGTGA
GACCCTGCGCCCTGGGGAGAAGGCCCCGTTAAGCAGAGAGCAACACCTCCAGTCTCCCAAGGAATGC
AGGACCTCTGCCAGTGATGTGGACGCTGACAACAACTGCCTAGGCACTGAGGTAGCTTAA

The disclosed MOL3a polypeptide (SEQ ID NO:8) encoded by SEQ ID NO:7 has 756 amino acid residues, and is presented using the one-letter code in Table 3B. The MOL3a protein was analyzed for signal peptide prediction and cellular localization. SignalP results predict that MOL3a is cleaved between position 31 and 32 (TTA-GG) of SEQ ID NO:8. Psort and Hydropathy profiles also predict that MOL3a is likely to be localized at the plasma membrane (certainty of 0.7300).

Table 3B. Encoded MOL3a protein sequence (SEQ ID NO:8).

MALPALGLDPWSLLGLFLFQLLQLLLPTTTAGGGGQGPMPRVRYYAGDERRALSFFHQKGLQDFDTLLLSGDGNT
LYVGAREAILALDIQDPGVPRLKNMIPWPASDRKKSECAFKKKSNETQCFNFIRVLVSYNVTHLYTCGTFAFSPA
CTFIELQDSYLLPISEDKVMEGKGQSPFDPAHKHTAVLVDGMLYSGTMNNFLGSEPILMRTLGSQPVLKTDNFLR
WLHHDASFVAAIPSTQVVYFFFEETASEFDFFERLHTSRVARVCKNDVGGEKLLQKKWTTFLKAQLLCTQPGQLP
FNVIRHAVLLPADSPTAPHIYAVFTSQWQVGGTRSSAVCAFSLLDIERVFKGKYKELNKETSRWTTYRGPETNPR
PGSCSVGPSSDKALTFMKDHFLMDEQVVGTPLLVKSGVEYTRLAVETAQGLDGKSHLVMYLGTSTGSLHKAVVSG
DSSAHLVEEIQLFPDPEPVRNLQLAPTQGAVFVGFSGGVWRVPRANCSVYESCVDCVLARDPHCAWDPESRLCSL
RNSWKQDMERGNPEWACASGPMSRSLRPQSRPQIVKEVLAVPNSILELPCPHLSALASYYWSHGPAAVPEASSTV
YNGSLLLIVQDGVGGLYQCWATENGFSYPVISYWVDSQDQTLALDPELAGIPREHVKVPLTRVSGGAALAAQQSY
WPHFVTVTVLFALVLSGALIILVASPLRALRARGKVQGCETLRPGEKAPLSREQHLQSPKECRTSASDVDADNNC
LGTEVA

The MOL3a nucleic acid sequence has 1398/1672 (83%) identical to a mouse Semaphorin B mRNA (GENBANK-ID: X85991).

The full amino acid sequence of MOL3a was found to have 628 of 760 (82%) identical to, 674 of 760 residues (88%) homologous with, the 760 amino acid residue Semaphorin B protein fr *mouse* (ptnr: SWISSNEW-ACC:Q62178).

MOL3a expression in different tissues was examined through TaqMan as described below in Example 1.

MOL3a also has high homology to the proteins disclosed in the BLASTP searches of the proprietary PATP database shown in Table 3C.

Table 3C. BLAST results for MOL3a

Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
patp:AAB24084	Human PRO1317 protein	761	751/761 (98%)	753/761 (98%)	0.0
patp:AAY99418	Human PRO1317 (UNQ783)	761	751/761 (98%)	753/761 (98%)	0.0
patp:AAB66043	Human TANGO 265	761	751/761 (98%)	753/761 (98%)	0.0
patp:AAB66167	Unidentified	761	751/761 (98%)	753/761 (98%)	0.0
patp:AAB37984	Human secreted protein encoded by gene 1 clone HTDAA93	762	743/761 (97%)	745/761 (97%)	0.0
patp:AAB66045	Human TANGO 265 mature protein	730	720/730 (98%)	722/730 (98%)	0.0
patp:AAB66046	Human TANGO 265 extracellular domain	652	642/652 (98%)	644/652 (98%)	0.0

Tissue Localization

MOL3a is expressed in at least the following tissues: Pituitary Gland, Thalamus

Chromosomal Localization

MOL3a maps to chromosome 1 .

MOL3b

In the present invention, the target sequence identified previously, MOL3a, was subjected to the exon linking process to confirm the sequence. PCR primers were designed by starting at the most upstream sequence available, for the forward primer, and at the most downstream sequence available for the reverse primer. In each case, the sequence was examined, walking inward from the respective termini toward the coding sequence, until a suitable sequence that is either unique or highly selective was encountered, or, in the case of the reverse primer, until the stop codon was reached. Such primers were designed based on *in silico* predictions for the full length cDNA, part (one or more exons) of the DNA or protein sequence of the target sequence, or by translated homology of the predicted exons to closely related human sequences sequences from other species. These primers were then employed in

PCR amplification based on the following pool of human cDNAs: adrenal gland, bone marrow, brain - amygdala, brain - cerebellum, brain - hippocampus, brain - substantia nigra, brain - thalamus, brain -whole, fetal brain, fetal kidney, fetal liver, fetal lung, heart, kidney, lymphoma - Raji, mammary gland, pancreas, pituitary gland, placenta, prostate, salivary gland, skeletal muscle, small intestine, spinal cord, spleen, stomach, testis, thyroid, trachea, uterus. Usually the resulting amplicons were gel purified, cloned and sequenced to high redundancy. The resulting sequences from all clones were assembled with themselves, with other fragments in CuraGen Corporation's database and with public ESTs. Fragments and ESTs were included as components for an assembly when the extent of their identity with another component of the assembly was at least 95% over 50 bp. In addition, sequence traces were evaluated manually and edited for corrections if appropriate. These procedures provide the sequence reported below, which is designated MOL3b, or alternatively Accession Number CG53027-02. This is a spliced variant of the previously identified sequence (Accession Number SC85516573_EXT) at amino acids 293-329.

A protein of the invention, referred to herein as MOL3b, is a human Semaphorin B-like protein. The novel nucleic acid of 2281 nucleotides (CG53027-02, SEQ ID NO:9) encoding a Semaphorin B-like protein is shown in Table 3D. An open reading frame (ORF) was identified beginning with a non-initiating codon for the mature protein at nucleotides 2-4 and ending with non-stop codon at nucleotides 2264-2266. The open reading frame may be extendable in both the 5' and 3' directions because of the lack of traditional start and stop codons. The nucleotide sequence is presented in Table 3D with the start and stop codons in bold letters and the 5' and 3' untranslated regions underlined.

Table 3D. MOL3b Nucleotide Sequence (SEQ ID NO:9)

GCCTGTGCCTAGAGTTTAAGCTACCTCAGTGCCTAGGCAGTTGTTGTCAGCGTCCACATCACTGGCAGAGGTCCT
GCATTCCTTGGGAGACTGGAGGTGTTGCTCTCTGCTTAACGGGGCCTTCTCCCCAGGGCGCAGGGTCTCACAGCC
CTGAACCTTGCCCCGAGCCCGGAGTGCTCTCAATGGGGAGGCCACGAGGATGATGAGGGCTCCTGAAAGCACTAA
GGCAAAGAGGACAGTGACAGTGACAAAGTGGGGCCAGTAGGACTGCTGGGCAGCCAGGGCGGCCCCACCACTGAC
CCTGGTCAACGGGACCTTCACATGCTCCCGGGGGATGCCTGCCAGTTCAGGATCCAGGGCCAGGGTCTGGTCCTG
GCTGTCCACCCAGTAGGAGATCACAGGGTATGAAAAGCCATTCTCAGTTGCCCAGCACTGGTAGAGACCCCCAAC
TCCATCCTGCACTATCAGCAAGAGGGAGCCATTGTAGACAGTGGAAGAGGCTTCTGGGACTGCTGCTGGGCCATG
ACTCCAATAATAAGAGGCCAAGGCTGACAGGTGGGGGCAGGGGAGCTCCAGGATGGAGTTGGGGACAGCCAGGAC
TTCTTTAATGATTTGCGGGCGGCTCTGAGGCCGAAGGCTCCTGCTCATGGGGCCACTGGCACATGCCCACTCTGG
GTTCCCCCGCTCCATGTCCTGCTTCCAGGAGTTCAGGTTGGGGGCAGACAGGAGGCAACAGGTTCGGGACTCAGG
GTCCCAGGCACAGTGGGGGTCCCGGGCAAGGACACAGTCCACACAGCTCTCATAGACACTACAGTTGGCTCGGGG
CACCCTCCAGACACCTCCTGAGAAGCCTACAAACACTGCACCCTGGGTGGGGGCCAGCTGCAGGTTGCGAACAGG
TTCAGGGTCAGGGAACAGCTGAATCTCTTCCACCAGATGAGCACTGCTGTCCCCACTTACCACAGCCTTGTGGAG
CGACCCTGTGGTGGTTCCCAGGTACATGACAAGATGGCTGTGCCCATCAAGGCCCTGGGCTGTCTCCACTGCAAG
CCGTGTATACTCCACGCCAGATTTCACCAGCAGGGGCGTCCCCACCACTTGCTCATCCATCAGGAAATGGTCCTT
CATGAAGGTCAGGGCCTTATCAGAGGAGGGGCCCACTGAGCAACTGCCTGGCCGGGGGTTGGTCTCAGGGCCCCT
ATAAGTAGTCCAGCGTGAAGTTTCTTTGTTCAACTCTTTGTATTTCCCCTTAAAGACACGTTCAATGTCCAAGAG
AGAGAAGGCACAAACCGCAGAGCTCCTGGTCCCGCCAACCTGCCACTGGGAGGTGAAGACTGCGTAGATGTGGGG
AGCTGTGGGAGAATCGGCGGGGAGCAGGACCGCGTGGCGGATGACGTTGAAGGGCAGCTGCCCCGGCTGGGTGCA
GAGCAGCTGGGCCTTCAGGAAGGTGGTCCACTTCTTCTGCAGCAGCTTTTCGCCGCCCACGTCATTCTTGCAGAC

```
TCTAGCCACCCGCGATGTGTGGAGCCTCTCAAAGAAGTCAAACTCGCTGGCTGTCTCCTCGAAGAAGAAGTAGAC
GACCTGGGTCGAAGGGATGGCTGCCACAAAGGAGGCGTCATGATGCAGCCAGCGGAGGAAGTTGTCGGTCTTGAG
GACAGGCTGGGATCCCAGTGTGCGCATCAGGATGGGCTCACTGCCCAGGAAGTTGTTCATAGTACCAGAATAGAG
CATCCCATCCACCAAGACAGCCGTATGCTTGTGAGCGGGGTCAAAGGGGCTTTGGCCTTTTCCCTCCATGACCTT
GTCCTCCGAGATGGGCAACAGGTAGGAATCTTGAAGTTCAATGAAGGTACAAGCAGGGCTGAAGGCGAAGGTGCC
GCAGGTGTAGAGATGGGTGACATTGTAAGAAACCAGGACACGGATGAAGTTGAAACACTGTGTCTCATTGCTCTT
CTTCTTAAAGGCACATTCACTCTTTTTTCTGTCACTGGCTGGCCACGGTATCATGTTCTTTAGCCTGGGGACCCC
TGGATCCTGGATATCCAAGGCCAGAATGGCTTCTCGAGCCCCCACGTAGAGAGTATTTCCATCACCACTCAGGAG
CAGAGTGTCAAAATCCTGGAGGCCCTTCTGGTGGAAGAAGCTAAGTGCCCTACGTTCATCCCCTGCATAGTATCT
GACCCTGGGCATGGGCCCCTGCCCGCCTCCCCCCGCGGTCGTCGTCGGCAGCAGCAGCTGAAGCAGTTGGAAGAG
GAAAAGGCCCAGGAGGCTCCAGGGGTCCAGG
```

The disclosed MOL3b polypeptide (SEQ ID NO:10) encoded by SEQ ID NO:9 has 754 amino acid residues, and is presented using the one-letter code in Table 3E. The MOL3b protein was analyzed for signal peptide prediction and cellular localization. SignalP results predict that MOL3b is cleaved between position 24 and 25 (TTA-GG) of SEQ ID NO:10. Psort and Hydropathy profiles also predict that MOL3b is likely to be localized at the plasma membrane (certainty of 0.7300).

Table 3E. Encoded MOL3b protein sequence (SEQ ID NO:10).

```
LDPWSLLGLFLFQLLQLLLPTTTAGGGGQGPMPRVRYYAGDERRALSFFHQKGLQDFDTLLLSGDGNTLYVGARE
AILALDIQDPGVPRLKNMIPWPASDRKKSECAFKKKSNETQCFNFIRVLVSYNVTHLYTCGTFAFSPACTFIELQ
DSYLLPISEDKVMEGKGQSPFDPAHKHTAVLVDGMLYSGTMNNFLGSEPILMRTLGSQPVLKTDNFLRWLHHDAS
FVAAIPSTQVVYFFFEETASEFDFFERLHTSRVARVCKNDVGGEKLLQKKWTTFLKAQLLCTQPGQLPFNVIRHA
VLLPADSPTAPHIYAVFTSQWQVGGTRSSAVCAFSLLDIERVFKGKYKELNKETSRWTTYRGPETNPRPGSCSVG
PSSDKALTFMKDHFLMDEQVVGTPLLVKSGVEYTRLAVETAQGLDGHSHLVMYLGTTTGSLHKAVVSGDSSAHLV
EEIQLFPDPEPVRNLQLAPTQGAVFVGFSGGVWRVPRANCSVYESCVDCVLARDPHCAWDPESRTCCLLSAPNLN
SWKQDMERGNPEWACASGPMSRSLRPQSRPQIIKEVLAVPNSILELPCPHLSALASYYWSHGPAAVPEASSTVYN
GSLLLIVQDGVGGLYQCWATENGFSYPVISYWVDSQDQTLALDPELAGIPREHVKVPLTRVSGGAALAAQQSYWP
HFVTVTVLPALVLSGALIILVASPLRALRARGKVQGCETLRPGEKAPLSREQHLQSPKECRTSASDVDADNNCLG
TEVA
```

The MOL3b nucleic acid sequence has 1910 of 2279 bases (83%) identical to a gb:GENBANK-ID:MMRNASEMB|acc:X85991.1 mRNA from *Mus musculus* (*M.musculus* mRNA for semaphorin B).

The full amino acid sequence of the protein of the invention was found to have 722 of 755 amino acid residues (95%) identical to, and 723 of 755 amino acid residues (95%) similar to, the 762 amino acid residue ptnr:TREMBLNEW-ACC:BAB20087 protein from Homo sapiens (Human) (SEMB).

The presence of identifiable domains in the protein disclosed herein was determined by searches versus domain databases such as Pfam, PROSITE, ProDom, Blocks or Prints and then identified by the Interpro domain accession number. Significant domains are summarized in Table 3F.

Table 3F Domain search for MOL3b

```
HMMER is freely distributed under the GNU General Public License (GPL).
```

```
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
HMM file:                 pfamHMMs
Sequence file:            /data4/genetools/kspytek35060Cg53027_01ProteinFasta.txt
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
Query:  CG53027_01

Scores for sequence family classification (score includes all domains):
Model         Description                                   Score    E-value  N
--------      -----------                                   -----    -------  ---
Sema          Sema domain                                   618.4   4.2e-182    1
Plexin_repeat Plexin repeat                                  22.0      0.013    1
integrin_B    Integrins, beta chain                           6.5      0.063    1

Parsed for domains:
Model         Domain  seq-f seq-t    hmm-f hmm-t      score  E-value
--------      ------- ----- -----    ----- -----      -----  -------
Sema            1/1      57   471 ..     1   490 []   618.4 4.2e-182
integrin_B      1/1     495   509 ..     1    14 [.     6.5    0.063
Plexin_repeat   1/1     489   555 ..     1    67 []    22.0    0.013
```

Tissue Localization

MOL3b is expressed in at least the following tissues: thalamus and Pituitary Gland. Expression information was derived from the tissue sources of the sequences that were included in the derivation of the sequence of MOL3b.

Chromosomal Localization

MOL3b maps to chromosome 1. This assignment was made using mapping information associated with genomic clones, public genes and ESTs sharing sequence identity with the disclosed sequence and CuraGen Corporation's Electronic Northern bioinformatic tool.

The disclosed MOL3a protein (SEQ ID NO:8) also has good identity with a number of other proteins, as shown in Table 3G.

Table 3G. BLAST results for MOL3a

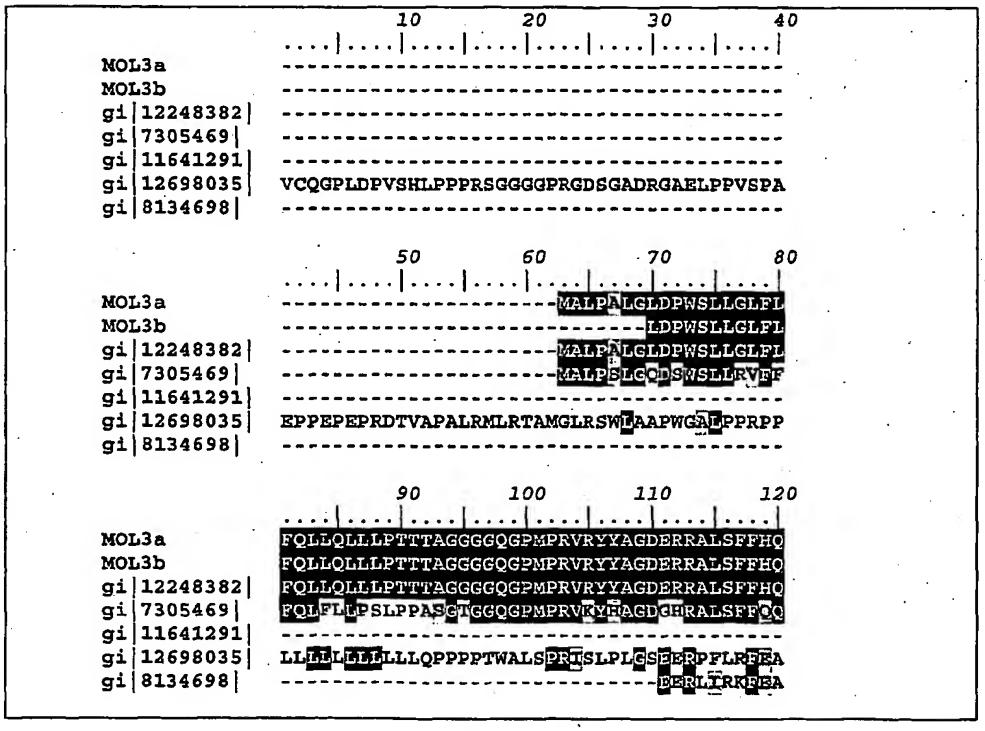
Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi|12248382|dbj|BAB20087.1| (AB029394)	SEMB [Homo sapiens]	762	719/762 (94%)	722/762 (94%)	0.0
gi|7305469|ref|NP_038686.1|	sema domain, immunoglobulin domain (Ig), transmembrane domain (TM) and short cytoplasmic domain, (semaphorin) 4A [Mus musculus]	760	626/760 (82%)	671/760 (87%)	0.0

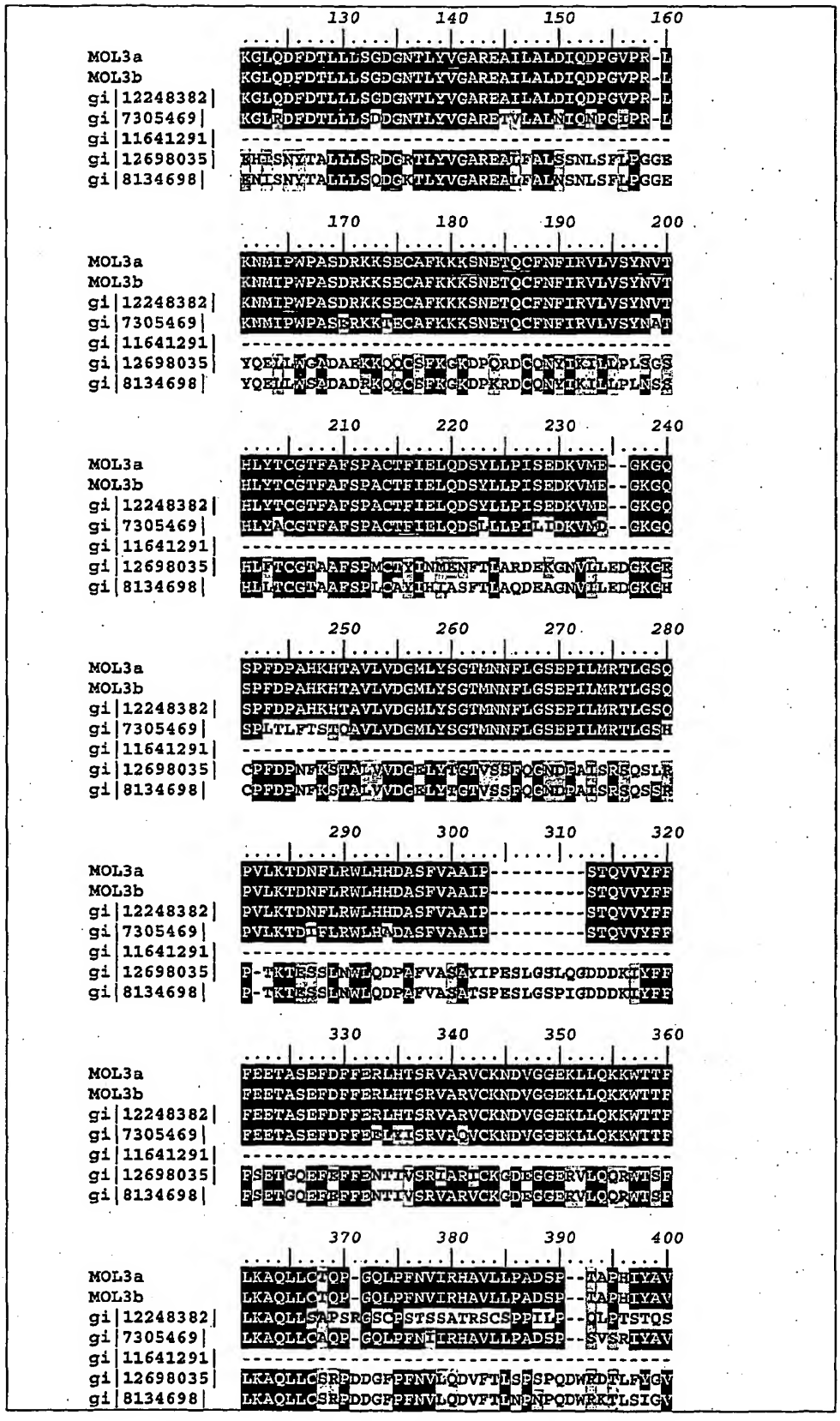
gi|11641291|ref|NP_071762.1|	hypothetical protein FLJ12287 similar to semaphorins [Homo sapiens]	328	318/328 (96%)	320/328 (96%)	0.0
gi|12698035|dbj|BAB21836.1| (AB051532)	KIAA1745 protein [Homo sapiens]	893	296/760 (38%)	422/760 (54%)	e-123
gi|8134698|sp|Q62179|SM4B_MOUSE	SEMAPHORIN 4B (SEMAPHORIN C) (SEMA C)	782	268/684 (39%)	382/684 (55%)	e-123

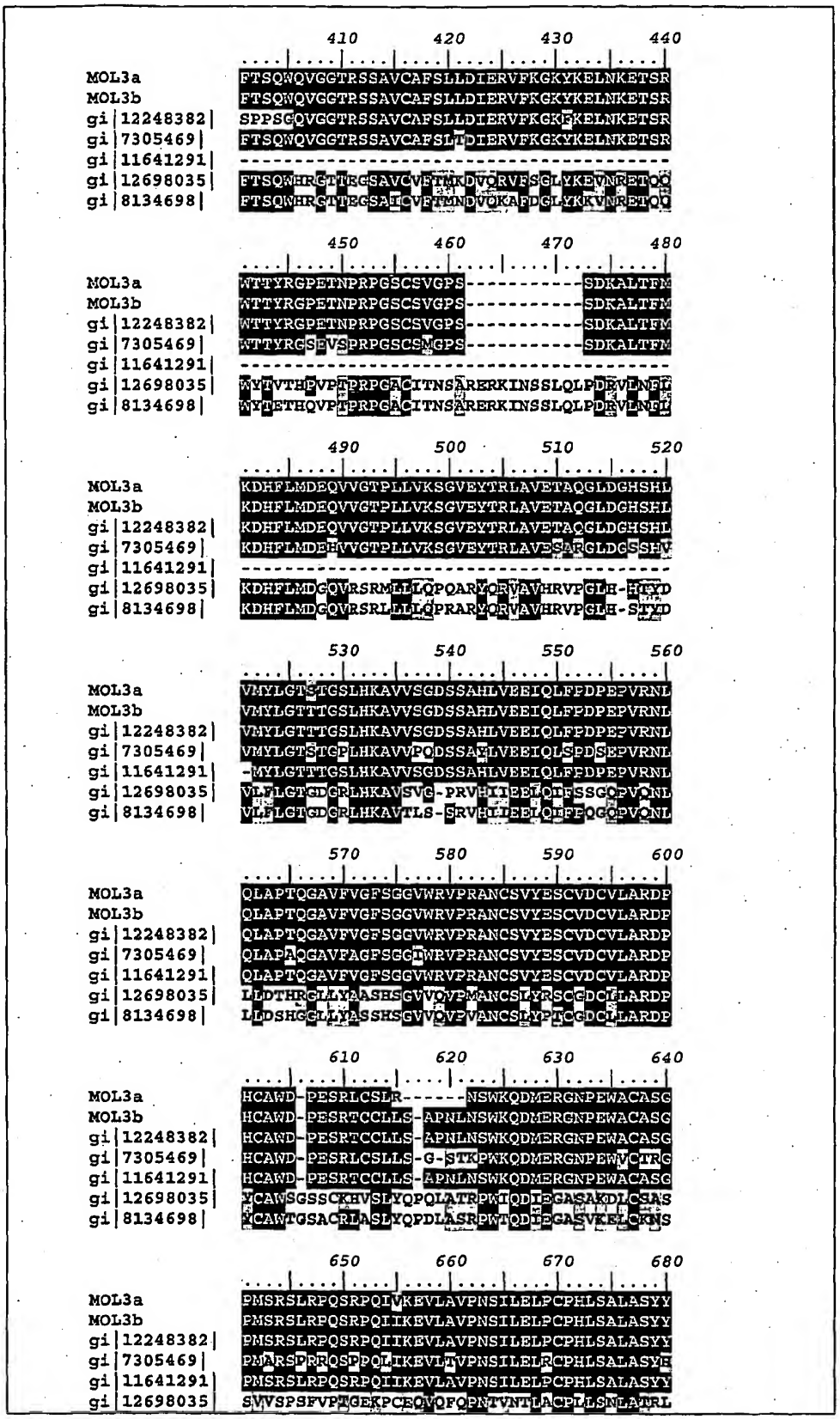
This information is presented graphically in the multiple sequence alignment given in Table 3H (with MOL3a being shown on line 1 and MOL3b on line 2) as a ClustalW analysis comparing MOL3 with related protein sequences.

Table 3H. Information for the ClustalW proteins:

1) MOL3a (SEQ ID NO:8)
2) MOL3b (SEQ ID NO:10)
3) gi|12248382|dbj|BAB20087.1| (AB029394) SEMB [Homo sapiens] (SEQ ID NO:35)
4) gi|7305469|ref|NP_038686.1| sema domain, immunoglobulin domain (Ig), transmembrane domain (TM)
and short cytoplasmic domain, (semaphorin) 4A [Mus musculus] (SEQ ID NO:36)
5) gi|11641291|ref|NP_071762.1| hypothetical protein FLJ12287 similar to semaphorins [Homo sapiens] (SEQ ID NO:37)
6) gi|12698035|dbj|BAB21836.1| (AB051532) KIAA1745 protein [Homo sapiens] (SEQ ID NO:38)
7) gi|8134698|sp|Q62179|SM4B_MOUSE SEMAPHORIN 4B (SEMAPHORIN C) (SEMA C) (SEQ ID NO:39)




130 140 150 160
MOL3a KGLQDFDTLLLSGDGNTLYVGAREAILALDIQDPGVPR-L
MOL3b KGLQDFDTLLLSGDGNTLYVGAREAILALDIQDPGVPR-L
gi|12248382| KGLQDFDTLLLSGDGNTLYVGAREAILALDIQDPGVPR-L
gi|7305469| KGLRDFDTLLLSDDGNTLYVGARETVLALNIQNPGEPR-L
gi|11641291| ----------
gi|12698035| EHESNYTALLLSRDGRTLYVGAREALFALSSNLSFLPGGE
gi|8134698| ENLSNYTALLLSQDGKTLYVGAREALFALNSNLSFLPGGE
170 180 190 200
MOL3a KNMIPWPASDRKKSECAFKKKSNETQCFNFIRVLVSYNVT
MOL3b KNMIPWPASDRKKSECAFKKKSNETQCFNFIRVLVSYNVT
gi|12248382| KNMIPWPASDRKKSECAFKKKSNETQCFNFIRVLVSYNVT
gi|7305469| KNMIPWPASERKKTECAFKKKSNETQCFNFIRVLVSYNAT
gi|11641291| ----------
gi|12698035| YQELLWGADAEKKQQCSFKGKDPQRDCQNYIKILLPLSGS
gi|8134698| YQELLWSQDADEKQQCSFKGKDPKRDCQNYIKILLPLNSS
210 220 230 240
MOL3a HLYTCGTFAFSPACTFIELQDSYLLPISEDKVME--GKGQ
MOL3b HLYTCGTFAFSPACTFIELQDSYLLPISEDKVME--GKGQ
gi|12248382| HLYTCGTFAFSPACTFIELQDSYLLPISEDKVME--GKGQ
gi|7305469| HLYACGTFAFSPACTFIELQDSLLLPILIDKVMD--GKGQ
gi|11641291| ----------
gi|12698035| HLFTCGTAAFSPMCTYINMENFTLARDEKGNVLLEDGKGK
gi|8134698| HLLTCGTAAFSPLCAYIHTASFTLAQDEAGNVLLEDGKGH
250 260 270 280
MOL3a SPFDPAHKHTAVLVDGMLYSGTMNNFLGSEPILMRTLGSQ
MOL3b SPFDPAHKHTAVLVDGMLYSGTMNNFLGSEPILMRTLGSQ
gi|12248382| SPFDPAHKHTAVLVDGMLYSGTMNNFLGSEPILMRTLGSQ
gi|7305469| SPLTLFTSTQAVLVDGMLYSGTMNNFLGSEPILMRTLGSH
gi|11641291| ----------
gi|12698035| CPFDPNFKSTALYVDGELYTGTVSSFQGNDPALSRSQSLR
gi|8134698| CPFDPNFKSTALYVDGELYTGTVSSFQGNDPALSRSQSSR
290 300 310 320
MOL3a PVLKTDNFLRWLHHDASFVAAIP---------STQVVYFF
MOL3b PVLKTDNFLRWLHHDASFVAAIP---------STQVVYFF
gi|12248382| PVLKTDNFLRWLHHDASFVAAIP---------STQVVYFF
gi|7305469| PVLKTDIFLRWLHHDASFVAAIP---------STQVVYFF
gi|11641291| ----------
gi|12698035| P-TKTESSLNWLQDPAFVASAYIPESLGSLQGDDDKLYFF
gi|8134698| P-TKTESSLNWLQDPAFVASATSPESLGSPIGDDDKLYFF
330 340 350 360
MOL3a FEETASEFDFFERLHTSRVARVCKNDVGGEKLLQKKWTTF
MOL3b FEETASEFDFFERLHTSRVARVCKNDVGGEKLLQKKWTTF
gi|12248382| FEETASEFDFFERLHTSRVARVCKNDVGGEKLLQKKWTTF
gi|7305469| FEETASEFDFFERLYISRVAQVCKNDVGGEKLLQKKWTTF
gi|11641291| ----------
gi|12698035| FSETGQEFEFFENTIVSRVARVCKGDEGGERVLQQRWTSF
gi|8134698| FSETGQEFEFFENTIVSRVARVCKGDEGGERVLQQRWTSF
370 380 390 400
MOL3a LKAQLLCTQP-GQLPFNVIRHAVLLPADSP--TAPHIYAV
MOL3b LKAQLLCTQP-GQLPFNVIRHAVLLPADSP--TAPHIYAV
gi|12248382| LKAQLLSAPSRGSCPSTSSATRSCSPPILP--QLPTSTQS
gi|7305469| LKAQLLCAQP-GQLPFNIIRHAVLLPADSP--SVSRIYAV
gi|11641291| ----------
gi|12698035| LKAQLLCSRPDDGFPFNVLQDVFTLSPSPQDWRDTLFYGV
gi|8134698| LKAQLLCSRPDDGFPFNVLQDVFTLNPNPQDWRKTLSIGV


410 420 430 440
MOL3a FTSQWQVGGTRSSAVCAFSLLDIERVFKGKYKELNKETSR
MOL3b FTSQWQVGGTRSSAVCAFSLLDIERVFKGKYKELNKETSR
gi|12248382| SPPSGQVGGTRSSAVCAFSLLDIERVFKGKYKELNKETSR
gi|7305469| FTSQWQVGGTRSSAVCAFSLLDIERVFKGKYKELNKETSR
gi|11641291| ----------
gi|12698035| FTSQWHRGTTRGSAVCVFTMKDVQRVFSGLYKEVNRETQQ
gi|8134698| FTSQWHRGTTRGSAICVFTMNDVQKAFDGLYKKVNRETQQ
450 460 470 480
MOL3a WTTYRGPETNPRPGSCSVGPS----------SDKALTFM
MOL3b WTTYRGPETNPRPGSCSVGPS----------SDKALTFM
gi|12248382| WTTYRGPETNPRPGSCSVGPS----------SDKALTFM
gi|7305469| WTTYRGSEVSPRPGSCSMGPS----------SDKALTFM
gi|11641291| ----------
gi|12698035| WYTVTHPVPTPRPGACITNSARERKINSSLQLPDEVLNET
gi|8134698| WYTETHQVPTPRPGACITNSARERKINSSLQLPDRVLNET
490 500 510 520
MOL3a KDHFLMDEQVVGTPLLVKSGVEYTRLAVETAQGLDGHSHL
MOL3b KDHFLMDEQVVGTPLLVKSGVEYTRLAVETAQGLDGHSHL
gi|12248382| KDHFLMDEQVVGTPLLVKSGVEYTRLAVETAQGLDGHSHL
gi|7305469| KDHFLMDEHVVGTPLLVKSGVEYTRLAVESARGLDGSSHV
gi|11641291| ----------
gi|12698035| KDHFLMDGQVRSRMLLLQPQARVQRVAVHRVPGLH-HTYD
gi|8134698| KDHFLMDGQVRSRLLLLQPRARVQRVAVHRVPGLH-STYD
530 540 550 560
MOL3a VMYLGTSTGSLHKAVVSGDSSAHLVEEIQLFFDPEPVRNL
MOL3b VMYLGTTTGSLHKAVVSGDSSAHLVEEIQLFFDPEPVRNL
gi|12248382| VMYLGTTTGSLHKAVVSGDSSAHLVEEIQLFFDPEPVRNL
gi|7305469| VMYLGTSTGPLHKAVVPQDSSAYLVEEIQLSPDSEPVRNL
gi|11641291| -MYLGTTTGSLHKAVVSGDSSAHLVEEIQLFFDPEPVRNL
gi|12698035| VLFLGTGDGRLHKAVSVG-PRVHIIEELQLFSSGQPVQNL
gi|8134698| VLFLGTGDGRLHKAVTLS-SRVHLIEELQLFFQGQPVQNL
570 580 590 600
MOL3a QLAPTQGAVFVGFSGGVWRVPRANCSVYESCVDCVLARDP
MOL3b QLAPTQGAVFVGFSGGVWRVPRANCSVYESCVDCVLARDP
gi|12248382| QLAPTQGAVFVGFSGGVWRVPRANCSVYESCVDCVLARDP
gi|7305469| QLAPAQGAVFAGFSGGIWRVPRANCSVYESCVDCVLARDP
gi|11641291| QLAPTQGAVFVGFSGGVWRVPRANCSVYESCVDCVLARDP
gi|12698035| LLDTHRGLLYAASHSGVVQVPMANCSLYRSCGDCLLARDP
gi|8134698| LLDSHGGLLYASSHSGVVQVPVANCSLYPICGDCLLARDP
610 620 630 640
MOL3a HCAWD-PESRLCSLR------NSWKQDMERGNPEWACASG
MOL3b HCAWD-PESRTCCLLS-APNLNSWKQDMERGNPEWACASG
gi|12248382| HCAWD-PESRTCCLLS-APNLNSWKQDMERGNPEWACASG
gi|7305469| HCAWD-PESRLCSLLS-G-STKPWKQDMERGNPEWVCTRG
gi|11641291| HCAWD-PESRTCCLLS-APNLNSWKQDMERGNPEWACASG
gi|12698035| YCAWSGSSCKHVSLYQPQLATRPWTQDIEGASAKDLCSAS
gi|8134698| FCAWTGSACRLASLYQPDLASRPWTQDIEGASVKELCKNS
650 660 670 680
MOL3a PMSRSLRPQSRPQIVKEVLAVPNSILELPCPHLSALASYY
MOL3b PMSRSLRPQSRPQIIKEVLAVPNSILELPCPHLSALASYY
gi|12248382| PMSRSLRPQSRPQIIKEVLAVPNSILELPCPHLSALASYY
gi|7305469| PMARSDRRQSPPQIIKEVLTVPNSILELRCPHLSALASYH
gi|11641291| PMSRSLRPQSRPQIIKEVLAVPNSILELPCPHLSALASYY
gi|12698035| SVVSPSFVPTGEKPCEQVQFQPNTVNTLACPLLSNLATRL

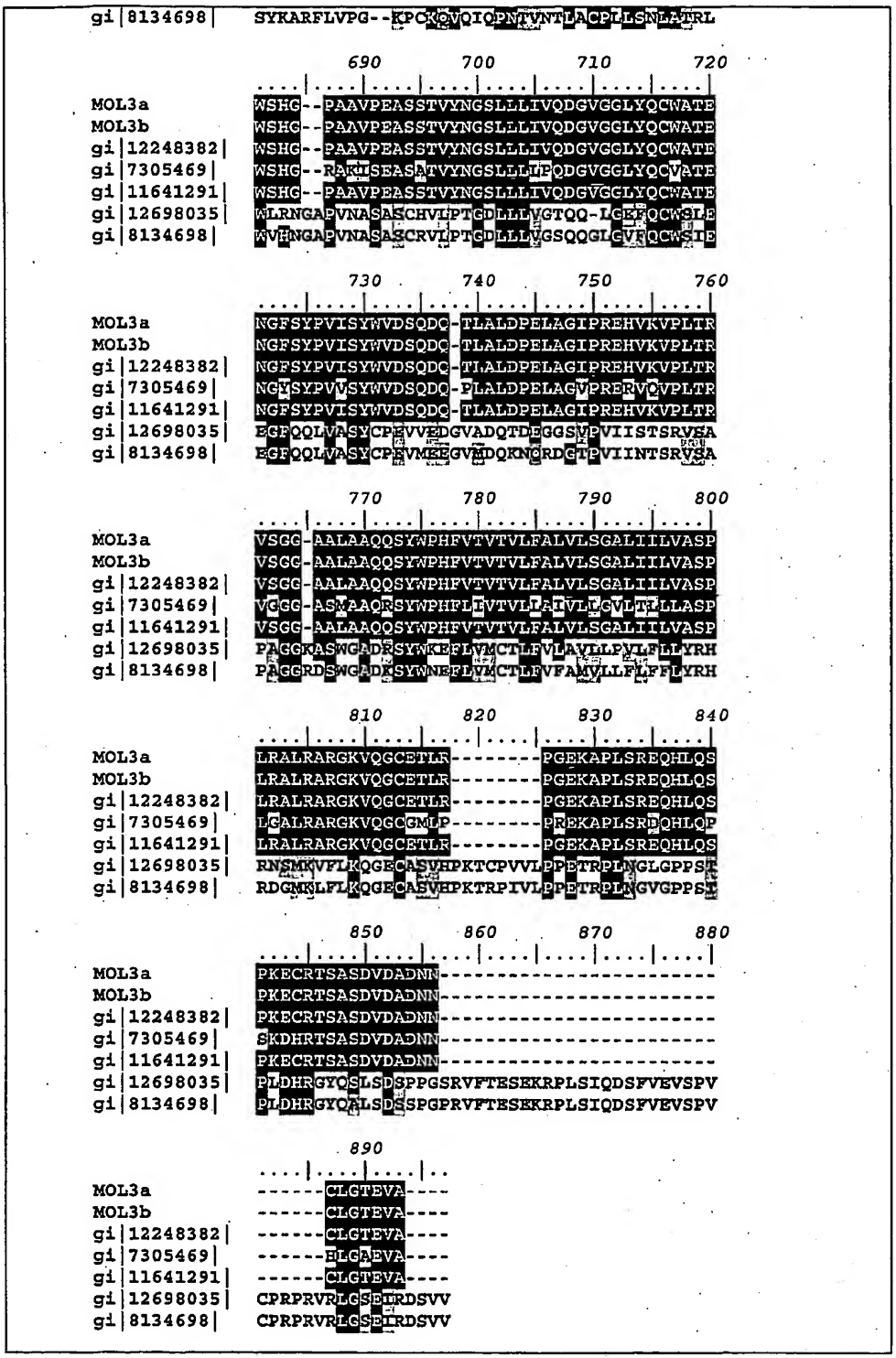


Table 3I lists the domain description from DOMAIN analysis results against MOL3. The region from amino acid residue 64 through 478 (SEQ ID NO:8) most probably ($E = 1e^{-121}$) contains a PSI, domain found in Plexins, Semaphorins and Integrins, aligned here in Table

3I. Semaphorins are involved in growth cone guidance as well as other developmental processes. Plexins and integrins are involved in developmental processes. The MOL1 sequence likely has properties similar to those of other proteins known to contain this domain

Table 3I. Domain Analysis of MOL3

gnl|Smart|smart00423, PSI, domain found in Plexins, Semaphorins and Integrins
CD-Length = 431 residues, 100.0% aligned
Score = 430 bits (1106), Expect = 1e-121

```
                                10        20        30        40        50
                        ....|....|....|....|....|....|....|....|....|....|
MOL3_3                  -NCSVY-ESCVDCVLARDPHCAWDPESRLCSLR-----------------
Smart|smart00423        RCSQY-TSCSECLLARDPYCGWCSSQGRCISG-----------------

                                  °°°  60        70
                                ....|....|....|....|....|..
MOL3_3                          -------NS---WKCDMERGNPEWACA
Smart|smart00423                ----ERCDS---SICNWSSG----GCP (SEQ ID NO:90)
```

The protein similarity information, expression pattern, cellular localization, and map location for the protein and nucleic acid for MOL3 suggest that this Semaphorin B-like protein may have important structural and/or physiological functions characteristic of the Semaphorin B family. This family is involved in developmental processes including growth cone guidance. MOL3 likely plays a similar role in those developmental processes. Therefore, the MOL3 nucleic acids and proteins are useful in potential diagnostic and therapeutic applications and as a research tool. These include serving as a specific or selective nucleic acid or protein diagnostic and/or prognostic marker, wherein the presence or amount of the nucleic acid or the protein are to be assessed. These also include potential therapeutic applications such as the following: (i) a protein therapeutic, (ii) a small molecule drug target, (iii) an antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), (iv) a nucleic acid useful in gene therapy (gene delivery/gene ablation), (v) an agent promoting tissue regeneration *in vitro* and *in vivo*, and (vi) a biological defense weapon.

The MOL3 nucleic acids and proteins have applications in the diagnosis and/or treatment of various diseases and disorders. For example, the compositions of the present invention will have efficacy for the treatment of patients suffering from: neuronal developmental, organizational, mediated and interactive disorders and disease; endocrine dysfunctions, diabetes, obesity, growth and reproductive disorders, injury repair as well as other diseases, disorders and conditions.

These materials are further useful in the generation of antibodies that bind immuno-specifically to the novel MOL3 substances for use in therapeutic or diagnostic methods. These

antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-MOLX Antibodies" section below. The disclosed MOL3 protein has multiple hydrophilic regions, each of which can be used as an immunogen. In one embodiment, a contemplated MOL3 epitope is from about amino acids 30 to 100. In another embodiment, a MOL3 epitope is from about amino acids 110 to 150. In additional embodiments, MOL3 epitopes are from about amino acids 160 to 200, 210 to 230, 250 to 300, 350 to 400, 450 to 475, 500 to 575, 620 to 630, and from about amino acids 700 to 750. These novel proteins can also be used to develop assay systems for functional analysis.

MOL4

The disclosed novel semaphorin-like protein, MOL4 (also referred to herein as SC_111750277_A), is encoded by a nucleic acid, 6408 nucleotides long (SEQ ID NO:11). An open reading frame was identified beginning with an ATG initiation codon at nucleotides 1400-1402 and ending with a TGA codon at nucleotides 5456-5458. Putative untranslated regions upstream from the initiation codon and downstream from the termination codon are underlined in Table 4A, and the start and stop codons are in bold letters.

Table 4A. MOL4 Nucleotide Sequence (SEQ ID NO:11).

CCTGGGACTCTGGGAGAATGGTCCAGAGCTCATTGTCCTTGTTGATAAAATGATAGATTTGGACTCAATATCCCA
TGCTGCCTCTTCCAACTTGATTTTTACCCCAGACTGGGCTACCAGACTGGTATGCCCACACATGCCCGTTTCCTT
TCTTTCTTCTCTGCATCTCTGCCTTTGTGTCCAGAGCGTGTTTTCCCTTTGCAAGTTTCTCTCCATTCTGCACA
TTATGAGTTTCAGCATTTCTGTTGCCCTAGAAAGTCTATCTTTGAGATCTTGCACTGTTTCTCTTTTTACAGTGT
CTCATAAACTCCCTTCTTGGATTCAGAACCACCCTTTCTTTCCCATTATCCTGTCAAACTGCTTCTTGCCATGGT
CCAGGGGTAGGAGGATGGCAGGCAGGAGGTGCTTCTCTGGGGCTCTTAGTGTCTCAATTCTTCTGCTTTATCTGG
GTTTTCCTTTACCCAGAATTTTATTATGTAAAATGCTTCACTCAGACTTTGTTCTAATTATCCAATTTTTGGCAT
ACTCTAGAAAGTCTTTTGATATTTTCCTTCCTCCAACTTATCTATTTTTATTTCATAGTTCTCTTTGGTTATCTC
TTAGAATCACACTTTCCTGGTTTTAATTTTTCAAATCCTTTGTCTTTCTCACTCGTTCTTAGGTCACCTTTTTTT
ACATTTTCAAATATATTTTTTGTTCAGCAGAGGGCTCCCTTCCCATCCCTCTTGCAGCCCGGGCAGCTAGGATTT
GAAGCTTGCCCCTTGAATCTTTCTCTCCCGCCTTCTAGCCATCAGAAACACTAGATCACTTAAACTTGTAAACAA
TTCGGCCTCGCTCCTTGTGATTGCGCTAAACCTTCCGTCCTCAGCTGAGAACGCTCCACCACCTCCCCGGATCGC
TCATCTCTTGGCTGCCCTCCCACTGTTCCTGATGTTATTTTACTCCCCGTATCCCCTACTCGTTCTTCACAATTC
TGTAGGGTGCGTATTACTAACCCCAGTTTACAGCTGAGGAAACTGAGGCTTGGAGAGGTTCGCTCGGTATCGTAC
AGTTTGCAAGGTTAACCCTAATCCGGCCAGTTCTGGCTTTCCAGCCCAGCCCAGCAGCCTAGCCTCCCTCTCTGC
CGCTGCAGGTTATAACGGCTCTCCCCCGTTTTACACGAGGTCCCTTCCCCTTCAAATCCACAGGCAGGAAGATCG
TTCCGAACTGACGGGGCTGGGGAATGTGGGAGTCCGGAGTGGGGTTTGGGGGAGCTTCCTCAGGCCCTGAGTGTT
GGGGTGGGCAGGCCGCGCCGATGGCCCTCGGGGATGTCACATTCGAGATGGGGTGACCGAGAACGGCAAGGCGGG
ATGTGGCAAACGGCGGCAAGTGCTCGGAGTCCTAGGTCTTGCCGCCGGAATGCCGGCCGGGGAAGGGGCTTCGGC
CCACCGGGCTGGTCACCACACTCGGCAGGCCCGGGGCGGGAGTCGGCCGAGCAGCCGCGGGATGCAGGGCGCCCC
CTCGCGCTCCTCCGCGCGCCTCGAGGCTGGCGGGTGCAGCGCCCGCCGCGGCAGGTCTGCTCCAGCCCCCTCCTC
TTTTTCGCTCCCGCTCCCCTCCTTCTCTCCCTTTGCTTGCAACTCCTCCCCCACCGCCCCCTCCCTCCTTCTGCT
CCCGCGGTCTCCTCCTCCCTGCTCTCTCCGAGCGCCGGGTCGGGAGCTAGTTGGAGCGCGGGGGTTGGTGCCAGA
GCCCAGCTCCGCCGAGCCGGGCGGGTCGGCAGCGCATCCAGCGGCTGCTGGGAGCCCGAGCGCAGCGGGCGCGGG
CCCGGGGTGGGGACTGCACCGGAGCGCTGAGAGCTGGAGGCCGTTCCTGCGCGGCCGCCCCATTCCCAGACCGGCC
GCCAGCCCATCTGGTTAGCTCCCGCCGCTCCGCGCCGCCCGGGAGTCGGGAGCCGCGGGGAACCGGGCACCTGCA
CCCGCCTCTGGGAGTGAGTGGTTCCAGCTGGTGCCTGGCCTGTGTCTCTTGGATGCCCTGTGGCTTCAGTCCGTC
TCCTGTTGCCCACCACCTCGTCCCTGGGCCGCCTGATACCCCAGCCCAACAGCTAAGGTGTGGATGGACAGTAGG
GGGCTGGCTTCTCTCACTGGTCAGGGGTCTTCTCCCCTGTCTGCCTCCCGGAGCTAGGACTGCAGAGGGGCCTAT
CATGGTGCTTGCAGGCCCCCTGGCTGTCTCGCTGTTGCTGCCCAGCCTCACACTGCTGGTGTCCCACCTCTCCAG
CTCCCAGGATGTCTCCAGTGAGCCCAGCAGTGAGCAGCAGCTGTGCGCCCTTAGCAAGCACCCCACCGTGGCCTT
TGAAGACCTGCAGCCGTGGGTCTCTAACTTCACCTACCCTGGAGCCCGGGATTTCTCCCAGCTGGCTTTGGACCC

CTCCGGGAACCAGCTCATCGTGGGAGCCAGGAACTACCTCTTCAGACTCAGCCTTGCCAATGTCTCTCTTCTTCA
GGCCACAGAGTGGGCCTCCAGTGAGGACACGCGCCGCTCCTGCCAAAGCAAAGGGAAGACTGAGGAGGAGTGTCA
GAACTACGTGCGAGTCCTGATCGTCGCCGGCCGGAAGGTGTTCATGTGTGGAACCAATGCCTTTTCCCCCATGTG
CACCAGCAGACAGGTGGGGAACCTCAGCCGGACTACTGAGAAGATCAATGGTGTGGCCCGCTGCCCCTATGACCC
ACGCCACAACTCCACAGCTGTCATCTCCTCCCAGGGGGAGCTCTATGCAGCCACGGTCATCGACTTCTCAGGTCG
GGACCCTGCCATCTACCGCAGCCTGGGCAGTGGGCCACCGCTTCGCACTGCCCAATATAACTCCAAGTGGCTTAA
TGAGCCAAACTTCGTGGCAGCCTATGATATTGGGCTGTTTGCATACTTCTTCCTGCGGGAGAACGCAGTGGAGCA
CGACTGTGGACGCACCGTGTACTCTCGCGTGGCCCGCGTGTGCAAGAATGACGTGGGGGGCCGATTCCTGCTGGA
GGACACATGGACCACATTCATGAAGGCCCGGCTCAACTGCTCCCGCCCGGGCGAGGTCCCCTTCTACTATAACGA
GCTGCAGAGTGCCTTCCACTTGCCAGAGCAGGACCTCATCTATGGAGTTTTCACAACCAACGTAAACAGCATCGC
GGCTTCTGCTGTCTGCGCCTTCAACCTCAGTGCTATCTCCCAGGCTTTCAATGGCCCATTTCGCTACCAGGAGAA
CCCCAGGGCTGCCTGGCTCCCCATAGCCAACCCCATCCCCAATTTCCAGTGTGGCACCCTGCCTGAGACCGGTCC
CAACGAGAACCTGACGGAGCGCAGCCTGCAGGACGCGCAGCGCCTCTTCCTGATGAGCGAGGCCGTGCAGCCGGT
GACACCCGAGCCCTGTGTCACCCAGGACAGCGTGCGCTTCTCACACCTCGTGGTGGACCTGGTGCAGGCTAAAGA
CACGCTCTACCATGTACTCTACATTGGCACCGAGTCGGGCACCATCCTGAAGGCGCTGTCCACGGCGAGCCGCAG
CCTCCACGGCTGCTACCTGGAGGAGCTGCACGTGCTGCCCCCCGGGCGCCGCGAGCCCCTGCGCAGCCTGCGCAT
CCTGCACAGCGCCCGCGCGCTCTTCGTGGGGCTGAGAGACGGCGTCCTGCGGGTCCCACTGGAGAGGTGCGCCGC
CTACCGCAGCCAGGGGGCATGCCTGGGGGCCCGGGACCCGTACTGTGGCTGGGACGGGAAGCAGCAACGTTGCAG
CACACTCGAGGACAGCTCCAACATGAGCCTCTGGACCCAGAACATCACCGCCTGTCCTGTGCGGAATGTGACACG
GGATGGGGGCTTCGGCCCATGGTCACCATGGCAACCATGTGAGCACTTGGATGGGGACAACTCAGGCTCTTGCCT
GTGTCGAGCTCGATCCTGTGATTCCCCTCGACCCCGCTGTGGGGGCCTTGACTGCCTGGGGCCAGCCATCCACAT
CGCCAACTGCTCCAGGAATGGGGCGTGGACCCCGTGGTCATCGTGGGCGCTGTGCAGCACGTCCTGTGGCATCGG
CTTCCAGGTCCGCCAGCGAAGTTGCAGCAACCCTGCTCCCCGCCACGGGGGCCGCATCTGCGTGGGCAAGAGCCG
GGAGGAACGGTTCTGTAATGAGAACACGCCTTGCCCGGTGCCCATCTTCTGGGCTTCCTGGGGCTCCTGGAGCAA
GTGCAGCAGCAACTGTGGAGGGGGCATGCAGTCGCGGCGTCGGGCCTGCGAGAACGGCAACTCCTGCCTGGGCTG
CGGCGTGGAGTTCAAGACGTGCAACCCCGAGGGCTGCCCCGAAGTGCGGCGCAACACCCCCTGGACGCCGTGGCT
GCCCGTGAACGTGACGCAGGGCGGGGCACGGCAGGAGCAGCGGTTCCGCTTCACCTGCCGCGCGCCCCTTGCAGA
CCCGCACGGCCTGCAGTTCGGCAGGAGAAGGACCGAGACGAGGACCTGTCCCGCGGACGGCTCCGGCTCCTGCGA
CACCGACGCCCTGGTGGAGGACCTCCTGCGCAGCGGGAGCACCTCCCCGCACACGGTGAGCGGGGGCTGGGCCGC
CTGGGGCCCGTGGTCGTCCTGCTCCCGGGACTGCGAGCTGGGCTTCCGCGTCCGCAAGAGAACGTGCACTAACCC
GGAGCCCCGCAACGGGGGCCTGCCCTGCGTGGGCGATGCTGCCGAGTACCAGGACTGCAACCCCCAGGCTTGCCC
AGTTCGGGGTGCTTGGTCCTGCTGGACCTCATGGTCTCCATGCTCAGCTTCCTGTGGTGGGGGTCACTATCAACG
CACCCGTTCCTGCACCAGCCCCGCACCCTCCCCAGGTGAGGACATCTGTCTCGGGCTGCACACGGAGGAGGCACT
ATGTGCCACACAGGCCTGCCCAGAAGGCTGGTCGCCCTGGTCTGAGTGGAGTAAGTGCACTGACGACGGAGCCCA
GAGCCGAAGCCGGCACTGTGAGGAGCTCCTCCCAGGGTCCAGCGCCTGTGCTGGAAACAGCAGCCAGAGCCGCCC
CTGCCCCTACAGCGAGATTCCCGTCATCCTGCCAGCCTCCAGCATGGAGGAGGCCACCGGCTGTGCAGGGTTCAA
TCTCATCCACTTGGTGGCCACGGGCATCTCCTGCTTCTTGGGCTCTGGGCTCCTGACCCTAGCAGTGTACCTGTC
TTGCCAGCACTGCCAGCGTCAGTCCCAGGAGTCCACACTGGTCCATCCTGCCACCCCCAACCATTTGCACTACAA
GGGCGGAGGCACCCCGAAGAATGAAAAGTACACACCCATGGAATTCAAGACCCTGAACAAGAATAACTTGATCCC
TGATGACAGAGCCAACTTCTACCCATTGCAGCAGACCAATGTGTACACGACTACTTACTACCCAAGCCCCCTGAA
CAAACACAGCTTCCGGCCCGAGGCCTCACCTGGACAACGGTGCTTCCCCAACAGCTGATACCGCCGTCCTGGGGA
CTTGGGCTTCTTGCCTTCATAAGGCACAGAGCAGATGGAGATGGGACAGTGGAGCCAGTTTGGTTTTCTCCCTCT
GCACTAGGCCAAGAACTTGCTGCCTTGCCTGTGGGGGGTCCCATCCGGCTTCAGAGAGCTCTGGCTGGCATTGAC
CATGGGGGAAAGGGCTGGTTTCAGGCTGACATATGGCCGCAGGTCCAGTTCAGCCCAGGTCTCTCATGGTTATCT
TCCAACCCACTGTCACGCTGACACTATGCTGCCATGCCTGGGCTGTGGACCTACTGGGCATTTGAGGAACTGGAG
AATGGAGATGGCAAGAGGGCAGGCTTTTAAGTTTGGGTTGGAGACAACTTCCTGTGGCCCCCACAAGCTGAGTCT
GGCCTTCTCCAGCTGGCCCCAAAAAAGGCCTTTGCTACATCCTGATTATCTCTGAAAGTAATCAATCAAGTGGCT
CCAGTAGCTCTGGATTTTCTGCCAGGGCTGGGCCATTGTGGTGCTGCCCCAGTATGACATGGGACCAAGGCCAGC
GCAGGTTATCCACCTCTGCCTGGAAGTCTATACTCTACCCAGGGCATCCCTCTGGTCAGAGGCAGTGAGTACTGG
GAACTGGAGGCTGACCTGTGCTTAGAAGTCCTTTAATCTGGGCTGGTACAGGCCTCAGCCTTGCCCTCAATGCAC
GAAAGGTGGCCCAGGAGAGAGGATCAATGCCACAGGAGGCAGAAGTCTGGCCTCTGTGCCTCTATGGAGACTATC
TTCCAGTTGCTGCTCAACAGAGTTGTTGGCTGAGACCTGCTTGGGAGTCTCTGCTGGCCCTTCATCTGTTCAGGA
ACACACACACACACACTCACACACGCACACACAATCACAATTTGCTACAGCAACAAAAAAGACATTGGGCTGT
GGCATTATTAATTAAAGATGATATCCAGTCTCC

The 1352 amino acid MOL4 polypeptide (SEQ ID NO:12) encoded by SEQ ID NO:11 is presented using the one-letter amino acid code in Table 4B. The Psort profile for MOL4 predicts that this sequence has no signal peptide and is likely to be localized in the plasma membrane with a certainty of 0.7900. MOL4 has a molecular weight of 145674.1 Daltons.

Table 4B. MOL4 protein sequence (SEQ ID NO:12)

```
MPAGEGASAHRAGHETRQARGGSRPSSRGMQGAPSRSSARLEAGGCSARRGRSAPAPSSFSLPLPSFSPFACNSSP
TAPSLLLLPRSPPPCSLRAPGRELVGARGLVPEPSSAEPGGSAAHPAAAGSPSAAGAGPGGDCTGALRAGGRSCAA
APFPDRPPAHLVSSRRSAPPGSREPRGTGHLHPPLGVSGSSWCLACVSWMPCGFSPSPVAHHLVPGPPDTPAQQLR
CGWTVGGWLLSLVRGLLPCLPPGARTAEGPIMVLAGPLAVSLLLPSLTLLVSHLSSSQDVSSEPSSEQQLCALSKH
PTVAFEDLQPWVSNFTYPGARDFSQLALDPSGNQLIVGARNYLFRLSLANVSLLQATEWASSEDTRRSCQSKGKTE
EECQNYVRVLIVAGRKVFMCGTNAFSPMCTSRQVGNLSRTTEKINGVARCPYDPRENSTAVISSQGELYAATVIDF
SGRDPAIYRSLGSGPPLRTAQYNSKWLNEPNFVAAYDIGLFAYFFLRENAVEHDCGRTVYSRVARVCKNDVGGRFL
LEDTWTTFMKARLNCSRPGEVPFYYNELQSAFHLPEQDLIYGVFTTNVNSIAASAVCAFNLSAISQAFNGPFRYQE
NPRAAWLPIANPIPNFQCGTLPETGPNENLTERSLQDAQRLFLMSEAVQPVTPEPCVTQDSVRFSHLVVDLVQAKD
TLYHVLYIGTESGTILKALSTASRSLHGCYLEELHVLPPGRREPLRSLRILHSARALFVGLRDGVLRVPLERCAAY
RSQGACLGARDPYCGWDGKQQRCSTLEDSSNMSLWTQNITACPVRNVTRDGGFGPWSPWQPCEHLDGDNSGSCLCR
ARSCDSPRPRCGGLDCLGPAIHIANCSRNGAWTPWSSWALCSTSCGIGFQVRQRSCSNPAPRHGGRICVGKSREER
FCNENTPCPVPIFWASWGSWSKCSSNCGGGMQSRRRACENGNSCLGCGVEFKTCNPEGCPEVRRNTPWTPWLPVNV
TQGGARQEQRFRFTCRAPLADPHGLQFGRRRTETRTCPADGSGSCDTDALVEDLLRSGSTSPHTVSGGWAAWGPWS
SCSRDCELGFRVRKRTCTNPEPRNGGLPCVGDAAEYQDCNPQACPVRGAWSCWTSWSPCSASCGGGHYQRTRSCTS
PAPSPGEDICLGLHTEEALCATQACPEGWSPWSEWSKCTDDGAQSRSRHCEELLPGSSACAGNSSQSRPCPYSEIP
VILPASSMEEATGCAGFNLIHLVATGISCFLGSGLLTLAVYLSCQHCQRQSQESTLVHPATPNHLHYKGGGTPKNE
KYTPMEFKTLNKNNLIPDDRANFYPLQQTNVYTTTYYPSPLNKHSFRPEASPGQRCFPNS
```

The disclosed nucleic acid MOL4 sequence has 3226 of 3664 bases (88%) identical to a *Mus musculus* semaphorin mRNA (GENBANK-ID: ACC: X97818).

The full amino acid sequence of the disclosed MOL4 polypeptide has 1021 of 1093 amino acid residues (93%) identical to, and 1053 of 1093 residues (96%) positive with, the 1093 amino acid residue semaphorin 5B precursor protein from *Mus musculus* (ptnr:SPTREMBL-ACC:Q60519), and 971 of 973 amino acid residues (99%) identical to, and 972 of 973 residues (99%) positive with patp:AAY94990 Human secreted protein vb21_1, having 999 aa. The C-terminal 1202 amino acid residues of MOL4 are 100% identical to human KIAA1445 PROTEIN (TREMBLNEW-ACC:BAA95969).

MOL4 expression in different tissues was examined through TaqMan as described below in Example 1.

A SNP for MOL4 and the corresponding amino acid change it would cause is shown in Table 4C. The SNP was identified using the techniques disclosed in Example 3.

Table 4C: SNP for MOL4

Consensus Position	Base Change	AA Change Position	Residue Change
2046	C > T	682	A > V

In a search of CuraGen's proprietary human expressed sequence assembly database, assemblies 111750277 (589 nucleotides) and 87739769 (896 nucleotides) were identified as having >95% homology to this predicted semaphorin sequence (Fig 3A2). This database is composed of the expressed sequences (as derived from isolated mRNA) from more than 96 different tissues. The mRNA is converted to cDNA and then sequenced. These expressed

DNA sequences are then pooled in a database and those exhibiting a defined level of homology are combined into a single assembly with a common consensus sequence. The consensus sequence is representative of all member components. Since the nucleic acid of the described invention has >95% sequence identity with the CuraGen assembly, the nucleic acid of the invention likely represents an expressed semaphorin sequence.

The DNA assembly 111750277 has 3 components and was found by CuraGen to be expressed in the following tissues: Lymph node and Lung. The DNA assembly 87739769 has 7 components and was found by CuraGen to be expressed in the following tissues: Brain, Uterus, and Lung.

BLASTP (Non-Redundant Composite database) analysis of the best hits for alignments with MOL4 are listed in Table 4D.

Table 4D. BLASTP results for MOL4

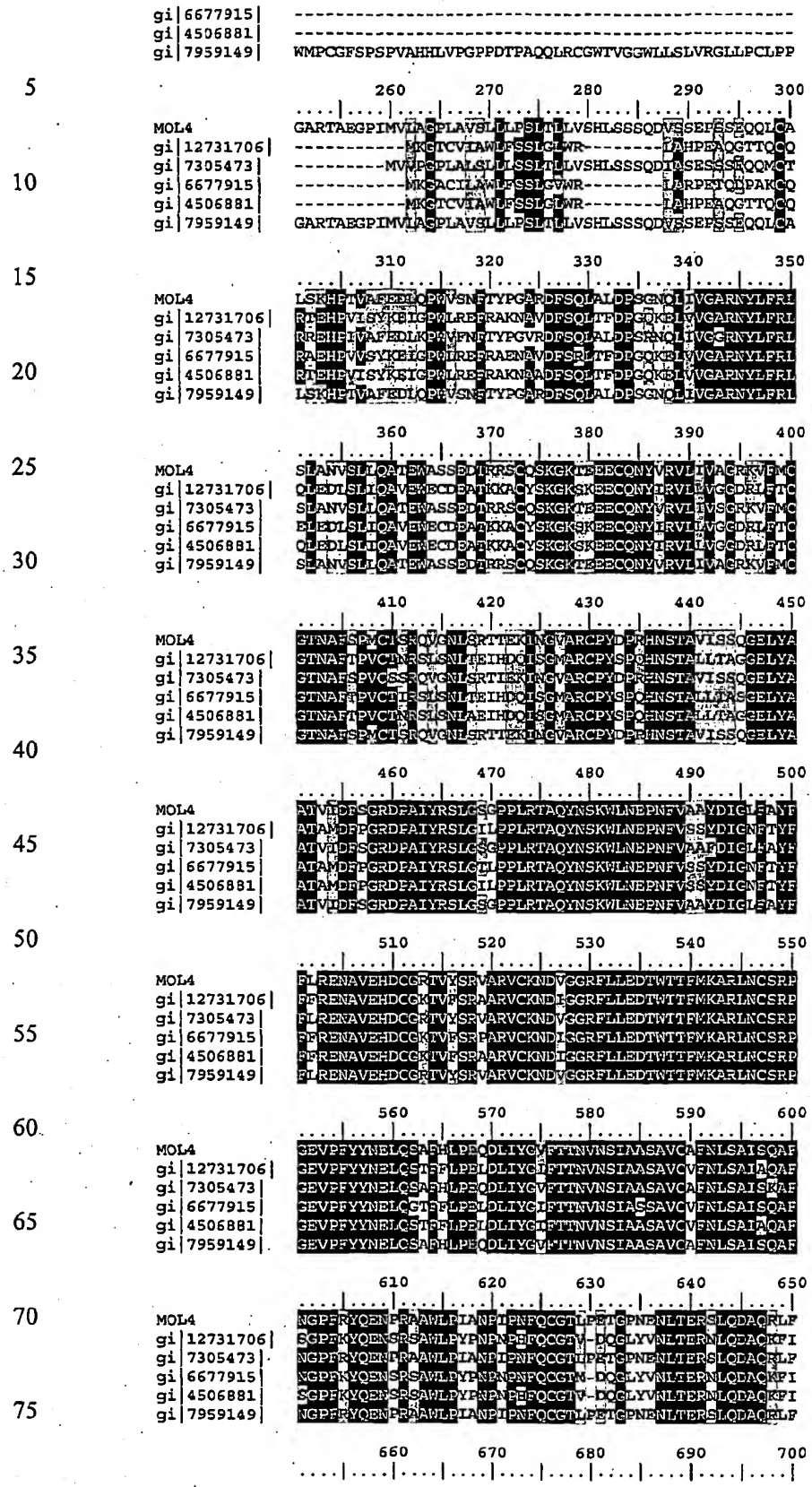
Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi|7305473|ref|NP_038689.1|	sema domain, seven thrombospondin repeats (type 1 and type 1-like), transmembrane domain (TM) and short cytoplasmic domain, (sem [Mus musculus]	1093	1021/1093 (93%)	1053/1093 (95%)	0.0
gi|7959149|dbj|BAA95969.1| (AB040878)	KIAA1445 protein [Homo sapiens]	1202	1202/1202 (100%)	1202/1202 (100%)	0.0
gi|4506881|ref|NP_003957.1|	sema domain, seven thrombospondin repeats (type 1 and type 1-like), transmembrane domain (TM) and short cytoplasmic domain, (semaphorin) 5A; semaphorin F; sema domain, seven thrombospondin repeats (type 1 and type 1-like), transmembrane domain (TM)	1074	616/1043 (59%)	781/1043 (74%)	0.0
gi|12731706|ref|XP_004042.2|	sema domain, seven thrombospondin repeats (type 1 and type 1-like), transmembrane domain (TM) and short cytoplasmic domain, (semaphorin) 5A [Homo sapiens]	1074	617/1043 (59%)	781/1043 (74%)	0.0
gi|6677915|ref|NP_033180.1|	sema domain, seven thrombospondin repeats (type 1 and type 1-like), transmembrane domain (TM) and short cytoplasmic domain, (sem; M-Sema D [Mus musculus]	1077	617/1046 (58%)	776/1046 (73%)	0.0

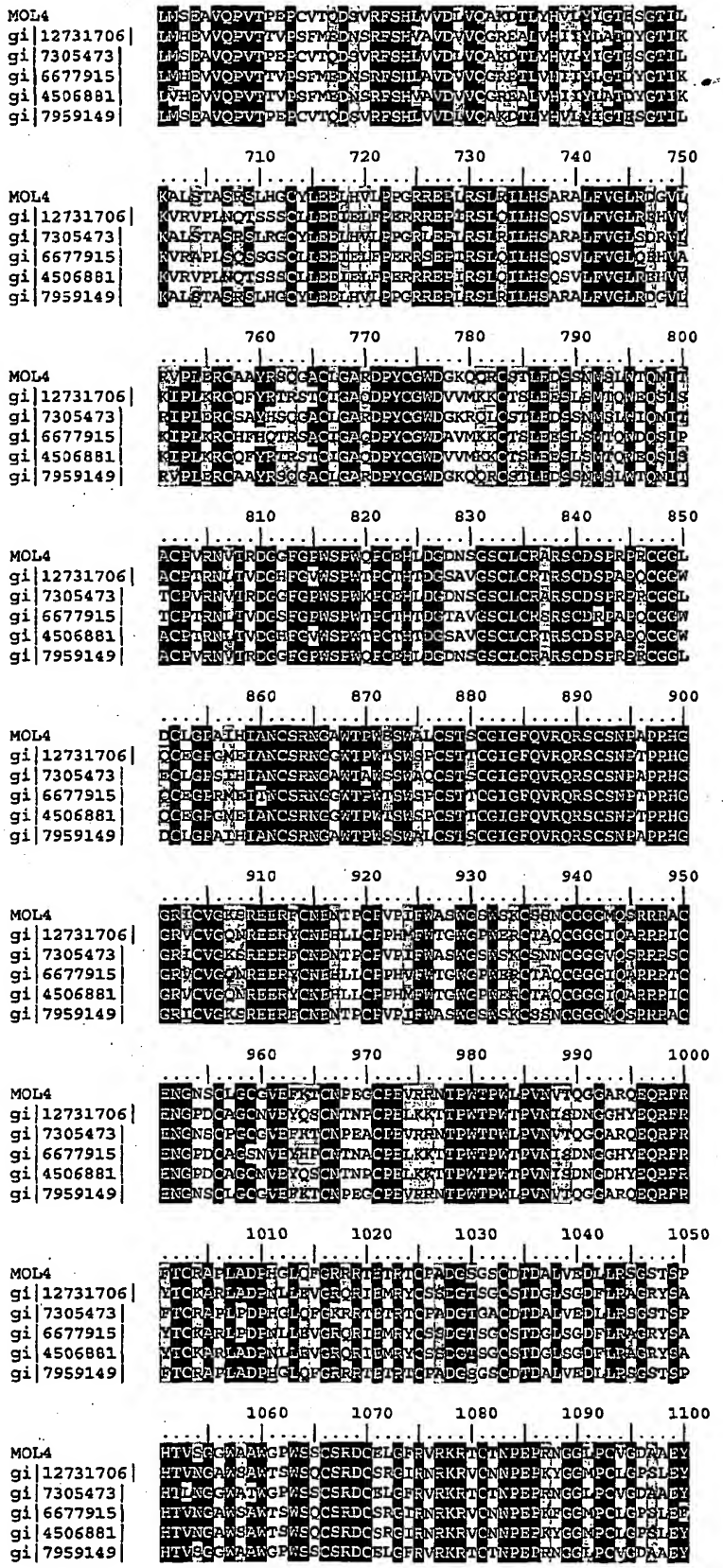
This information is presented graphically in the multiple sequence alignment given in Table 4E (with MOL4 being shown on line 1) as a ClustalW analysis comparing MOL4 with related sequences.

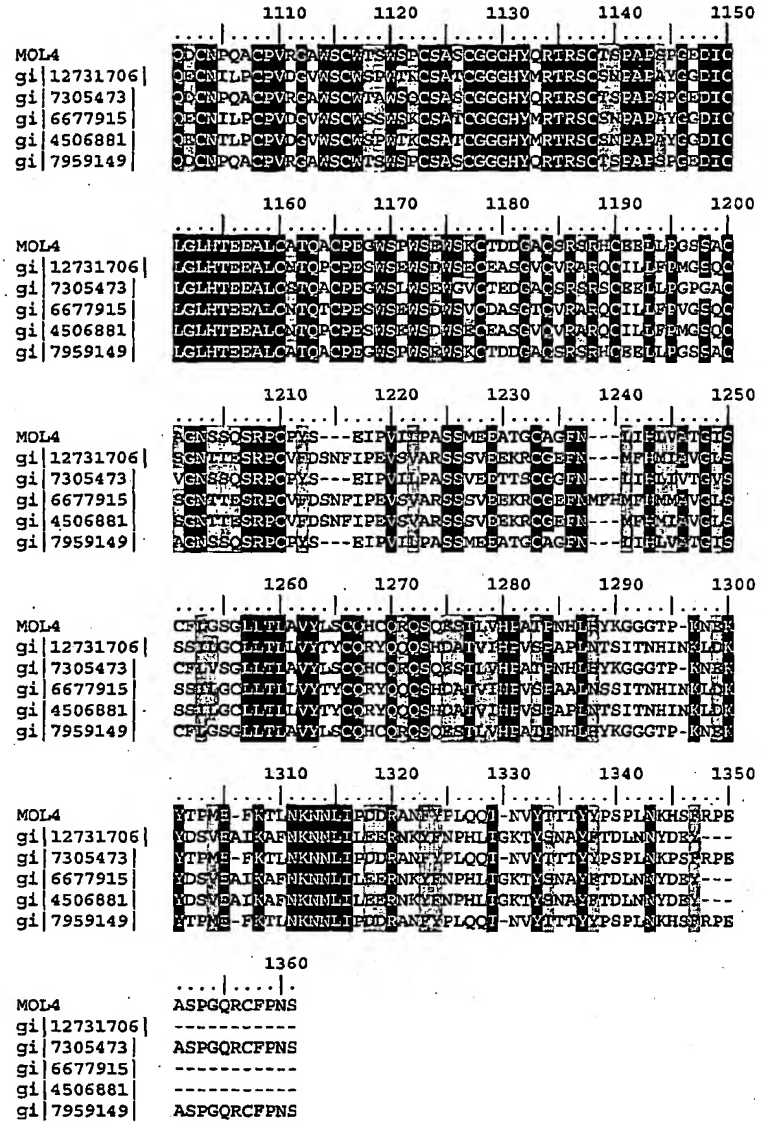
Table 4E Information for the ClustalW proteins:

1) MOL4 (SEQ ID NO:12)
2) gi|12731706|ref|XP_004042.2| sema domain, seven thrombospondin repeats (type 1 and type 1-like), transmembrane domain (TM) and short cytoplasmic domain, (semaphorin) 5A [Homo sapiens] (SEQ ID NO:41)
3) gi|7305473|ref|NP_038689.1| sema domain, seven thrombospondin repeats (type 1 and type 1-like), transmembrane domain (TM) and short cytoplasmic domain, (sem [Mus musculus] (SEQ ID NO:42)
4) gi|6677915|ref|NP_033180.1| sema domain, seven thrombospondin repeats (type 1 and type 1-like), transmembrane domain (TM) and short cytoplasmic domain, (sem; M-Sema D [Mus musculus] (SEQ ID NO:43)
5) gi|4506881|ref|NP_003957.1| sema domain, seven thrombospondin repeats (type 1 and type 1-like), transmembrane domain (TM) and short cytoplasmic domain, (semaphorin) 5A; semaphorin F; sema domain, seven thrombospondin repeats (type 1 and type 1-like), transmembrane domain (TM) (SEQ ID NO:44)
6) gi|7959149|dbj|BAA95969.1| (AB040878) KIAA1445 protein [Homo sapiens] (SEQ ID NO:45)

```
                    10        20        30        40        50
                    ....|....|....|....|....|....|....|....|....|....|
MOL4                MPAGEGASAHRAGHHTRQARGGSRPSSRGMQGAPSRSSARLEAGGCSARR
gi|12731706|        --------------------------------------------------
gi|7305473|         --------------------------------------------------
gi|6677915|         --------------------------------------------------
gi|4506881|         --------------------------------------------------
gi|7959149|         --------------------------------------------------

                    60        70        80        90        100
                    ....|....|....|....|....|....|....|....|....|....|
MOL4                GRSAPAPSSFSLPLPSFSPFACNSSPTAPSLLLLPRSPPPCSLRAPGREL
gi|12731706|        --------------------------------------------------
gi|7305473|         --------------------------------------------------
gi|6677915|         --------------------------------------------------
gi|4506881|         --------------------------------------------------
gi|7959149|         --------------------------------------------------

                    110       120       130       140       150
                    ....|....|....|....|....|....|....|....|....|....|
MOL4                VGARGLVPEPSSAEPGGSAAHPAAAGSPSAAGAGPGGDCTGALRAGGRSC
gi|12731706|        --------------------------------------------------
gi|7305473|         --------------------------------------------------
gi|6677915|         --------------------------------------------------
gi|4506881|         --------------------------------------------------
gi|7959149|         --------------------------------------------------

                    160       170       180       190       200
                    ....|....|....|....|....|....|....|....|....|....|
MOL4                AAAPFPDRPPAHLVSSRRSAPPGSREPRGTGHLHPPLGVSGSSWCLACVS
gi|12731706|        --------------------------------------------------
gi|7305473|         --------------------------------------------------
gi|6677915|         --------------------------------------------------
gi|4506881|         --------------------------------------------------
gi|7959149|         AAAPFPDRPPAHLVSSRRSAPPGSREPRGTGHLHPPLGVSGSSWCLACVS

                    210       220       230       240       250
                    ....|....|....|....|....|....|....|....|....|....|
MOL4                WMPCGFSPSPVAHHLVPGPPDTPAQQLRCGWTVGGWLLSLVRGLLPCLPP
gi|12731706|        --------------------------------------------------
gi|7305473|         --------------------------------------------------
```


gi|6677915|
gi|4506881|
gi|7959149|
MOL4
gi|12731706|
gi|7305473|
gi|6677915|
gi|4506881|
gi|7959149|
260 270 280 290 300
310 320 330 340 350
360 370 380 390 400
410 420 430 440 450
460 470 480 490 500
510 520 530 540 550
560 570 580 590 600
610 620 630 640 650
660 670 680 690 700


MOL4
gi|12731706|
gi|7305473|
gi|6677915|
gi|4506881|
gi|7959149|
710 720 730 740 750
760 770 780 790 800
810 820 830 840 850
860 870 880 890 900
910 920 930 940 950
960 970 980 990 1000
1010 1020 1030 1040 1050
1060 1070 1080 1090 1100



Tables 4F-4K list the domain descriptions from DOMAIN analysis results against MOL4. The region from amino acid residue 327 through 725 (SEQ ID NO:12) most probably ($E = 2e^{-118}$) contains a Sema domain found in Semaphorins, aligned here in Table 4F. Semaphorins are involved in growth cone guidance, axonal pathfinding, and other developmental processes. The region from amino acid residue 1057 through 1109 (SEQ ID NO:12) most probably ($E = 3e^{-9}$) contains a Thrombospondin type-1 repeat found in thrombospondin-1 that binds to and activates TGF-beta, aligned here in Table 4G. TGF-beta is involved in the modulation of proliferation in many cell types. The region from amino acid residue 868 through 921 (SEQ ID NO:12) most probably ($E = 4e^{-8}$) also contains a

Thrombospondin type-1 repeat found in thrombospondin-1 that binds to and activates TGF-beta, aligned here in Table 4H. The region from amino acid residue 926 through 972 (SEQ ID NO:12) most probably ($E = 6e^{-7}$) also contains a Thrombospondin type-1 repeat found in thrombospondin-1 that binds to and activates TGF-beta, aligned here in Table 4I. The region from amino acid residue 1169 through 1210 (SEQ ID NO:12) most probably ($E = 0.001$) also contains a Thrombospondin type-1 repeat found in thrombospondin-1 that binds to and activates TGF-beta, aligned here in Table 4J. The region from amino acid residue 756 through 803 (SEQ ID NO:12) most probably ($E = 1e^{-5}$) also contains a Thrombospondin type-1 repeat found in thrombospondin-1 that binds to and activates TGF-beta, aligned here in Table 4K. The presence of these domains indicates that the MOL4 sequence has properties similar to those of other proteins known to contain these domains.

Table 4F. Domain Analysis of MOL4

gnl|Smart|smart00630, Sema, semaphorin domain
CD-Length = 430 residues, 99.8% aligned
Score = 420 bits (1080), Expect = 2e-118

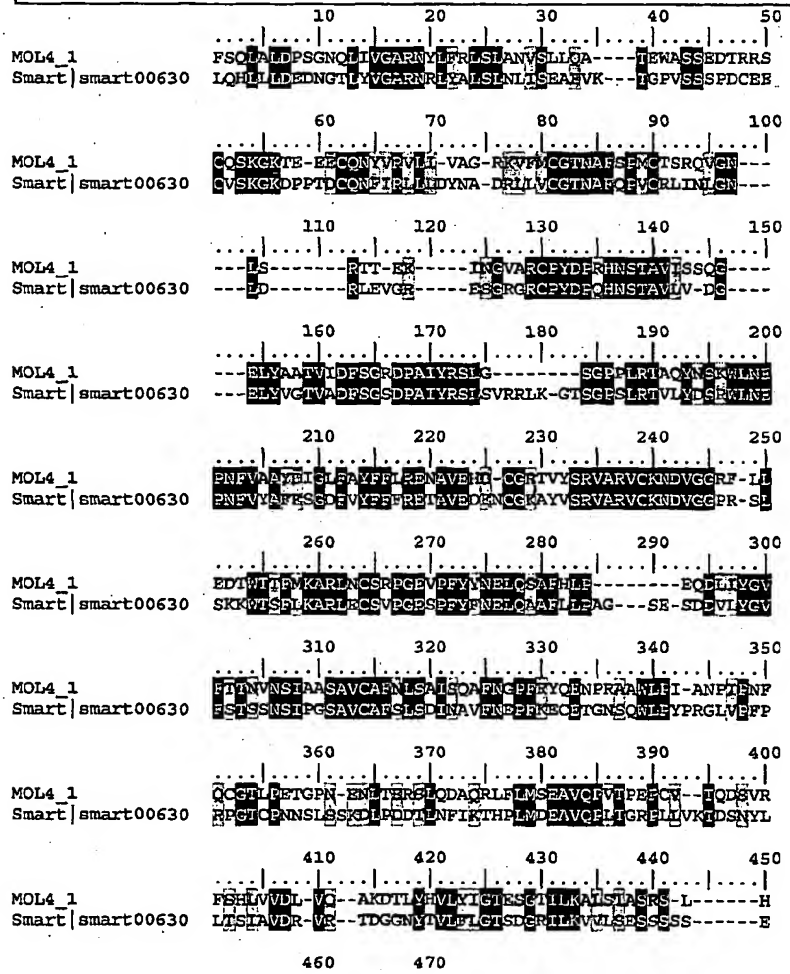


Table 4G. Domain Analysis of MOL4

gnl|Smart|smart00209, TSP1, Thrombospondin type 1 repeats; Type 1 repeats in thrombospondin-1 bind and activate TGF-beta.
CD-Length = 51 residues, 100.0% aligned
Score = 58.2 bits (139), Expect = 3e-09

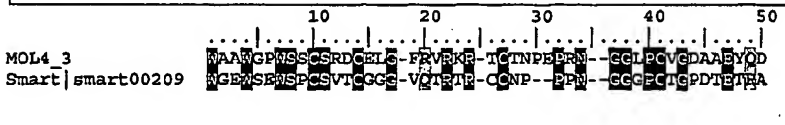


Table 4H. Domain Analysis of MOL4

gnl|Smart|smart00209, TSP1, Thrombospondin type 1 repeats; Type 1 repeats in thrombospondin-1 bind and activate TGF-beta.
CD-Length = 51 residues, 100.0% aligned
Score = 54.3 bits (129), Expect = 4e-08

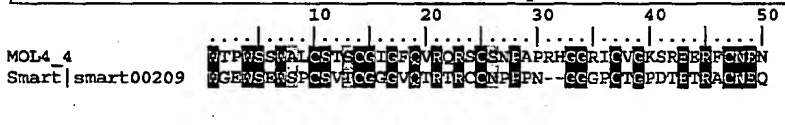


Table 4I. Domain Analysis of MOL4

gnl|Smart|smart00209, TSP1, Thrombospondin type 1 repeats; Type 1 repeats in thrombospondin-1 bind and activate TGF-beta.
CD-Length = 51 residues, 100.0% aligned
Score = 50.4 bits (119), Expect = 6e-07

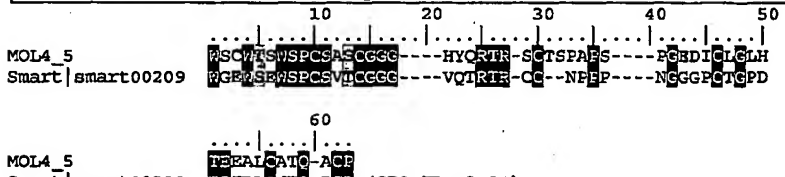


Table 4J. Domain Analysis of MOL4

gnl|Smart|smart00209, TSP1, Thrombospondin type 1 repeats; Type 1 repeats in thrombospondin-1 bind and activate TGF-beta.
CD-Length = 51 residues, 88.2% aligned
Score = 39.7 bits (91), Expect = 0.001

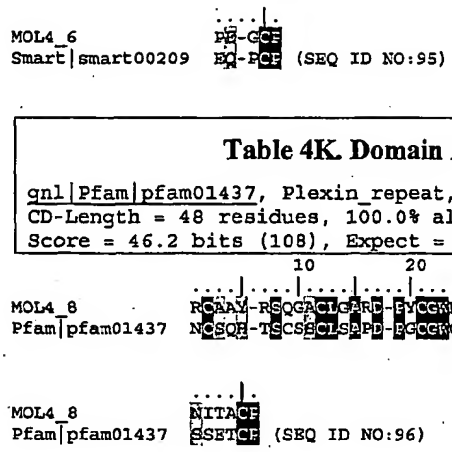


Table 4K. Domain Analysis of MOL4

gnl|Pfam|pfam01437, Plexin_repeat, Plexin repeat
CD-Length = 48 residues, 100.0% aligned
Score = 46.2 bits (108), Expect = 1e-05

The above defined information for MOL4 suggests that this semaphorin-like protein may function as a member of a "Semaphorin family". Therefore, the novel nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for MOL4 include, but are not limited to: protein therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene therapy (gene delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

The MOL4 nucleic acids and proteins are useful in potential therapeutic applications implicated in Parkinson's disease, psychotic and neurological disorders, Alzheimers disease, cancer including but not limited to lung or breast cancer, endocrine disorders, inflammatory disorders, gastro-intestinal disorders and disorders of the respiratory system, and/or other pathologies and disorders. For example, a cDNA encoding the semaphorin-like protein may be useful in gene therapy, and the semaphorin-like protein may be useful when administered to a subject in need thereof. By way of nonlimiting example, the compositions of the present invention will have efficacy for treatment of patients suffering from Parkinson's disease, psychotic and neurological disorders, Alzheimers disease, cancer including but not limited to lung or breast cancer, endocrine disorders, inflammatory disorders, gastro-intestinal disorders and disorders of the respiratory system. MOL4, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

These materials are further useful in the generation of antibodies that bind immuno-specifically to the novel MOL4 substances for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from

hydrophobicity charts, as described in the "Anti-MOLX Antibodies" section below. The disclosed MOL4 protein has multiple hydrophilic regions, each of which can be used as an immunogen. In one embodiment, a contemplated MOL4 epitope is from about amino acids 5 to 75. In another embodiment, a MOL4 epitope is from about amino acids 100 to 200. In additional embodiments, MOL4 epitopes are from about amino acids 300 to 375, 420 to 600, 600 to 675, 775 to 850, 900 to 1150, , and from about amino acids 1250 to 1350. These novel proteins can also be used to develop assay systems for functional analysis.

MOL5

MOL5a

The disclosed novel semaphorin 4C -like nucleic acid of 3868 nucleotides, MOL5a, (also referred to as SC20422974-A) is shown in Table 5A. An ORF begins with an ATG initiation codon at nucleotides 453-455 and ends with a TGA codon at nucleotides 2952-2954. A putative untranslated region upstream from the initiation codon and downstream from the termination codon is underlined in Table 5A, and the start and stop codons are in bold letters.

Table 5A. MOL5a Nucleotide Sequence (SEQ ID NO:13)

CGACTATCCATGAAGCCCGGAGCCCCAGTGGCTGCAAGGCCTGCTGCCTGAGGTTCTTTCAAGAAACTCAAACCT
CTTAGGCCTGAGTGTGTATGTTGGGCGGGGGTCCCCTTTTTATTTCTCAAATGATTTCCTGTTGCGCAGAGGTAG
TGGTGGGTCTGGAGGCCAGGGAGGGCTTCCCGGAGCCTGTTTAGCCTTCAGCCAACTCAACTCCTCCCCGCTTCC
CAGGGAGACCTGTGGTCTTTTAGGCAGAGGCCAAGTGTGGGGACTTAGGTCCACCTCCAAAGAGAAGGGGAAGGA
GGGCACCGGGGCTCCTGGAAGGCCTGATGAGGAGTCCTGTGGCCTCTCCTGCTGCGGGCCCCTCTGGTTTGCTTT
CTCTGGCTGTGATTTCTGACCATGTCTTTTCCCTCAGCAGGACAGCTGGCCTGAAGCTCAGAGCCGGGGCGTGCG
CCATGGCCCCACACTGGGCTGTCTGGCTGCTGGCAGCAAGGCTGTGGGGCCTGGGCATTGGGGCTGAGGTGTGGT
GGAACCTTGTGCCGCGTAAGACAGTGTCTTCTGGGGAGCTGGCCACGGTAGTACGGCGGTTCTCCCAGACCGGCA
TCCAGGACTTCCTGACACTGACGCTGACGGAGCCCACTGGGCTTCTGTACGTGGGCGCCCGAGAGGCGCTGTTTG
CCTTCAGTGTAGAGGCTCTGGAGCTGCAAGGAGCGATCTCCTGGGAGGCCCCCGTGGAGAAGAAGACTGAGTGTA
TCCAGAAAGGGAAGAACAACCAGACCGAGTGCTTCAACTTCATCCGCTTCCTGCAGCCCTACAATGCCTCCCACC
TGTACGTCTGTGGCACCTACGCCTTCCAGCCCAAGTGCACCTACGTCAACATGCTCACCTTCACTTTGGAGCATG
GAGAGTTTGAAGATGGGAAGGGCAAGTGTCCCTATGACCCAGCTAAGGGCCATGCTGGCCTTCTTGTGGATGGTG
AGCTGTACTCGGCCACACTCAACAACTTCCTGGGCACGGAACCCATTATCCTGCGTAACATGGGGCCCCACCACT
CCATGAAGACAGAGTACCTGGCCTTTTGGCTCAACGAACCTCACTTTGTAGGCTCTGCCTATGTACCTGAGAGTG
TGGGCAGCTTCACGGGGGACGACGACAAGGTCTACTTCCTCTTCAGGGAGCGGGCAGTGGAGTCCGCCTGCTATG
CCGAGCAGGTGGTGGCTCGTGTGGCCCGTGTCTGCAAGGGCGATATGGGGGGCGCACGGACCCTGCAGAGGAAGT
GGACCACGTTCCTGAAGGCGCGGCTGGCATGCTCTGCCCCGAACTGGCAGCTCTACTTCAACCAGCTGCAGGCGA
TGCACACCCTGCAGGACACCTCCTGGCACAACACCACCTTCTTTGGGGTTTTTCAAGCACAGTGGGGTGACATGT
ACCTGTCGGCCATCTGTGAGTACCAGTTGGAAGAGATCCAGCGGGTGTTTGAGGGCCCCTATAAGGAGTACCATG
AGGAAGCCCAGAAGTGGGACCGCTACACTGACCCTGTACCCAGCCCTCGGCCTGGCTCGTGCATTAACAACTGGC
ATCGGCGCCACGGCTACACCAGCTCCCTGGAGCTACCCGACAACATCCTCAACTTCGTCAAGAAGCACCCGCTGA
TGGAGGAGCAGGTGGGGCCTCGGTGGAGCCGCCCCCTGCTCGTGAAGAAGGGCACCAACTTCACCCACCTGGTGG
CCGACCGGGTTACAGGACTTGATGGAGCCACCTATACAGTGCTGTTCATTGGCACAGGTCAGGCATGGCTGCTCA
AGGCTGTGAGCCTGGGGCCCTGGGTTCACCTGATTGAGGAGCTGCAGCTGTTTGACCAGGAGCCCATGAGAAGCC
TGGTGCTATCTCAGTCGCAGAAGCTGCTCTTTGCCGGCTCCCGCTCTCAGCTGGTGCAGCTGCCCGTGGCCGACT
GCATGAAGTATCGCTCCTGTGCAGACTGTGTCCTCGCCCGGGACCCCTATTGCGCCTGGAGCGTCAACACCAGCC
GCTGTGTGGCCGTGGGTGGCCACTCTGGGTCCTTTCTGATCCAGCATGTGATGACCTCGGACACTTCAGGCATCT
GCAACCTCCGTGGCAGTAAGAAAGTCAGGCCCACTCCCAAAAACATCACGGTGGTGGCGGGCACAGACCTGGTGC
TGCCCTGCCACCTCTCCTCCAACTTGGCCCATGCCCGCTGGACCTTTGGGGGCCGGGACCTGCCTGCGGAACAGC
CCGGGTCCTTCCTCTACGATGCCCGGCTCCAGGCCCTGGTTGTGATGGCTGCCCAGCCCCGCCATGCCGGGGCCT
ACCACTGCTTTTCAGAGGAGCAGGGGGCGCGGCTGGCTGCTGAAGGCTACCTTGTGGCTGTCGTGGCAGGCCCGT
CGGTGACCTTGGAGGCCCGGGCCCCCCTGGAAAACCTGGGGCTGGTGTGGCTGGCGGTGGTGGCCCTGGGGGCTG

TGTGCCTGGTGCTGCTGCTGCTGGTGCTGTCATTGCGCCGGCGGCTGCGGGAAGAGCTGGAGAAAGGGGCCAAGG
CTACTGAGAGGACCTTGGTGTACCCCCTGGAGCTGCCCAAGGAGCCCACCAGTCCCCCCTTCCGGCCCTGTCCTG
AACCAGATGAGAAACTTTGGGATCCTGTCGGTTACTACTATTCAGATGGCTCCCTTAAGATAGTACCTGGGCATG
CCCGGTGCCAGCCCGGTGGGGGGCCCCCTTCGCCACCTCCAGGCATCCCAGGCCAGCCTCTGCCTTCTCCAACTC
GGCTTCACCTGGGGGGTGGGCGGAACTCAAATGCCAATGGTTACGTGCGCTTACAACTAGGAGGGGAGGACCGGG
GAGGGCTCGGGCACCCCCTGCCTGAGCTCGCGGATGAACTGAGACGCAAACTGCAGCAACGCCAGCCACTGCCCG
ACTCCAACCCCGAGGAGTCATCAGTATGAGGGGAACCCCCACCGCGTCGGCGGGAAGCGTGGGAGGTGTAGCTCC
TACTTTGCACAGGCACCAGCTATCTCAGGGACATGGCACGGGCACCTGCTCTGTCTGGGACAGATACTGCCCAG
CACCCACCCGGCCATGAGGACCTGCTCTGCTCAGCACGGGCACTGCCACTTGGTGTGGCTCACCAGGGCACCAGC
CTCGCAGAAGGCATCTTCCTCCTCTCTGTGAATCACAGACACGCGGGACCCCAGCCGCCAAAACTTTTCAAGGCA
GAAGTTTCAAGATGTGTGTTTGTCTGTATTTGCACATGTGTTTGTGTGTGTGTGTATGTGTGTGTGCACGCGCGT
GCGCGCTTGTGGCATAGCTTCCTGTTTCTGTCAAGTCTTCCCTTGGCCTGGGTCCTCCTGGTGAGTCATTGGAGC
TATGAAGGGGAAGGGGTCGTATCACTTTGTCTCTCCTACCCCCACTGCCCCGAGTGTCGGGCAGCGATGTACATA
TGGAGGTGGGGTGGACAGGGTGCTGTGCCCCTTCAGAGGGAGTGCAGGGCTTGGGGTGGGCCTAGTCCTGCTCCT
AGGGCTGTGAATGTTTTCAGGGTGGGGGGAGGGAGATGGAGCCTCCTGTGTGTTTGGGGGGAAGGGTGGGTGGGG
CCTCCCACTTGGCCCCGGGGTTCAGTGGTATTTTATACTTGCCTTCTTCCTGTACAGGGCTGGGAAAGGCTGTGT
GAGGGGAGAGAAGGGAGAGGGTGGGCCTGCTGTGGACAATGGCATACTCTCTTCCAGCCCTAGGAGGAGGGCTCC
TAACAGTGTAACTTATTGTGTCCCCGCGTATTTATTTGTTGTAAATATTTGAGTATTTTTATATTGACAAATAAA
ATGGAGAAAAAAAAAAAAAAAAAAAAAAAAAGTCGTATCGATGT

The MOL5a protein encoded by SEQ ID NO:13 has 833 amino acid residues and is presented using the one-letter code in Table 5B. The Psort profile for MOL5a predicts that this sequence is likely to be localized at the mitochondrial inner membrane with a certainty of 0.8000 or plasma membrane with a certainty of 0.7000. MOL5a has a cleavage site between amino acods 20 and 21 (GIG-AE), and a molecular weight of 92617.0 Daltons.

Table 5B. Encoded MOL5a protein sequence (SEQ ID NO:14)

MAPHWAVWLLAARLWGLGIGAEVWWNLVPRKTVSSGELATVVRRFSQTGIQDFLTLTLTEPTGLLYVGAREAL
FAFSVEALELQGAISWEAPVEKKTECIQKGKNNQTECFNFIRFLQPYNASHLYVCGTYAFQPKCTYVNMLTFT
LEHGEFEDGKGKCPYDPAKGHAGLLVDGELYSATLNNFLGTEPIILRNMGPHHSMKTEYLAFWLNEPHFVGSA
YVPESVGSFTGDDDKVYFLFRERAVESDCYAEQVVARVARVCKGDMGGARTLQRKWTTFLKARLACSAPNWQL
YFNQLQAMHTLQDTSWHNTTFFGVFQAQWGDMYLSAICEYQLEEIQRVFEGPYKEYHEEAQKWDRYTDPVPSP
RPGSCINNWHRRHGYTSSLELPDNILNFVKKHPLMEEQVGPRWSRPLLVKKGTNFTHLVADRVTGLDGATYTV
LFIGTGQAWLLKAVSLGPWVHLIEELQLFDQEPMRSLVLSQSQKLLFAGSRSQLVQLPVADCMKYRSCADCVL
ARDPYCAWSVNTSRCVAVGGHSGSFLIQHVMTSDTSGICNLRGSKKVRPTPKNITVVAGTDLVLPCHLSSNLA
HARWTFGGRDLPAEQPGSFLYDARLQALVVMAAQPRHAGAYHCFSEEQGARLAAEGYLVAVVAGPSVTLEARA
PLENLGLVWLAVVALGAVCLVLLLLVLSLRRRLREELEKGAKATERTLVYPLELPKEPTSPPFRPCPEPDEKL
WDPVGYYYSDGSLKIVPGHARCQPGGGPPSPPPGIPGQPLPSPTRLHLGGGRNSNANGYVRLQLGGEDRGGLG
HPLPELADELRRKLQQRQPLPDSNPEESSV

The disclosed nucleic acid sequence for MOL5a has 2917 of 3443 bases (84%) identical to a semaphorin 4C mRNA (GENBANK-ID: S79463|acc:S79463) (E= 0.0).

The full MOL5a amino acid sequence has 729 of 834 amino acid residues (87%) identical to, and 772 of 834 residues (92%) positives with, the 834 amino acid semaphorin 4C Precursor protein from *Mus musculus* (Mouse) (ptnr:SPTREMBL-ACC: Q64151) (E= 0.0). In addition, this protein contains the following protein domains (as defined by Interpro) at the indicated nucleotide positions: Sema domain (a.a. 53-481; IPR001627), integrin_B (a.a. 505-519; IPR000413), Plexin_repeat (a.a. 499-551; IPR002165), ig (a.a. 570-629; IPR000353)

MOL5a expression in different tissues was examined through TaqMan as described below in Example 1.

Chromosomal Localization

MOL5a has been localized to human chromosome 2.

MOL5b

Another disclosed novel semaphorin 4C -like nucleic acid of 2558 nucleotides, MOL5b, (also referred to as SC14998905_EXT) is shown in Table 5C. An ORF begins with an ATG initiation codon at nucleotides 21 - 23 and ends with a TGA codon at nucleotides 2520-2522. A putative untranslated region upstream from the initiation codon and downstream from the termination codon is underlined in Table 5C, and the start and stop codons are in bold letters.

Table 5C. MOL5b Nucleotide Sequence (SEQ ID NO:15)

TCAGAGCCGGGGCGTGCGCC**ATG**GCCCCACACTGGGCTGTCTGGCTGCTGGCAGCAAGGCTGTGGGGCCTGGGCA
TTGGGGCTGAGGTGTGGTGGAACCTTGTGCCGCGTAAGACAGTGTCTTCTGGGGAGCTGGCCACGGTAGTACGGC
GGTTCTCCCAGACCGGCATCCAGGACTTCCTGACACTGACGCTGACGGAGCCCACTGGGCTTCTGTACGTGGGCG
CCAGGGACCATGCCTCTGCACTGGGCGTCCCTGTGTTGCTGCTGCAGGCTGTGATCTCCTGGGAGGCCCCCGTGG
AGAAGAAGACTGAGTGTATCCAGAAAGGGAAGAACAACCAGACCGAGTGCTTCAACTTCATCCGCTTCCTGCAGC
CCTACAATGCCTCCCACCTGTACGTCTGTGGCACCTACGCCTTCCAGCCCAAGTGCACCTACGTCAACATGCTCA
CCTTCACTTTGGAGCATGGAGAGTTTGAAGATGGGAAGGGCAAGTGTCCCTATGACCCAGCTAAGGGCCATGCTG
GCCTTCTTGTGGATGGTGAGCTGTACTCGGCCACACTCAACAACTTCCTGGGCACGGAACCCATTATCCTGCGTA
ACATGGGGCCCCACCACTCCATGAAGACAGAGTACCTGGCCTTTTGGCTCAACGAACCTCACTTTGTAGGCTCTG
CCTATGTACCTGAGAGTGTGGGCAGCTTCACGGGGGACGACGACAAGGTCTACTTCTTCTTCAGGGAGCGGGCAG
TGGAGTCCGACTGCTATGCCGAGCAGGTGGTGGCTCGTGTGGCCCGTGTCTGCAAGGGCGATATGGGGGGCGCAC
GGACCCTGCAGAGGAAGTGGACCACGTTCCTGAAGGCGCGGCTGGCATGCTCTGCCCCGAACTGGCAGCTCTACT
TCAACCAGCTGCAGGCGATGCACACCCTGCAGGACACCTCCTGGCACAACACCACCTTCTTTGGGGTTTTTCAAG
CACAGTGGGGTGACATGTACCTGTCGGCCATCTGTGAGTACCAGTTGGAAGAGATCCAGCGGGTGTTTGAGGGCC
CCTATAAGGAGTACCATGAGGAAGCCCAGAAGTGGGACCGCTACACTGACCCTGTACCCAGCCCTCGGCCTGGCT
CGTGCATTAACAACTGGCATCGGCGCCACGGCTACACCAGCTCCCTGGAGCTACCCGACAACATCCTCAACTTCG
TCAAGAAGCACCCGCTGATGGAGGAGCAGGTGGGGCCTCGGTGGAGCCGCCCCCTGCTCGTGAAGAAGGGCACCA
ACTTCACCCACCTGGTGGCCGACCGGGTTACAGGACTTGATGGAGCCACCTATACAGTGCTGTTCATTGGCACAG
GAGACGGCTGGCTGCTCAAGGCTGTGAGCCTGGGGCCCTGGGTTCACCTGATTGAGGAGCTGCAGCTGTTTGACC
AGGAGCCCATGAGAAGCCTGGTGCTATCTCAGAGCAAGAAGCTGCTCTTTGCCGGCTCCCGCTCTCAGCTGGTGC
AGCTGCCCGTGGCCGACTGCATGAAGTATCGCTCCTGTGCAGACTGTGTCCTCGCCCGGGACCCCTATTGCGCCT
GGAGCGTCAACACCAGCCGCTGTGTGGCCGTGGGTGGCCACTCTGGATCTCTACTGATCCAGCATGTGATGACCT
CGGACACTTCAGGCATCTGCAACCTCCGTGGCAGTAAGAAAGTCAGGCCCACTCCCAAAAACATCACGGTGGTGG
CGGGCACAGACCTGGTGCTGCCCTGCCACCTCTCCTCCAACTTGGCCCATGCCCGCTGGACCTTTGGGGGCCGGG
ACCTGCCTGCGGAACAGCCCGGGTCCTTCCTCTACGATGCCCGGCTCCAGGCCCTGGTTGTGATGGCTGCCCAGC
CCCGCCATGCCGGGGCCTACCACTGCTTTTCAGAGGAGCAGGGGGCGCGGCTGGCTGCTGAAGGCTACCTTGTGG
CTGTCGTGGCAGGCCCGTCGGTGACCTTGGAGGCCCGGGCCCCCCTGGAAAACCTGGGGCTGGTGTGGCTGGCGG
TGGTGGCCCTGGGGGCTGTGTGCCTGGTGCTGCTGCTGCTGGTGCTGTCATTGCGCCGGCGGCTGCGGGAAGAGC
TGGAGAAAGGGGCCAAGGCTACTGAGAGGACCTTGGTGTACCCCCTGGAGCTGCCCAAGGAGCCCACCAGTCCCC
CCTTCCGGCCCTGTCCTGAACCAGATGAGAAACTTTGGGATCCTGTCGGTTACTACTATTCAGATGGCTCCCTTA
AGATAGTACCTGGGCATGCCCGGTGCCAGCCCGGTGGGGGGCCCCCTTCGCCACCTCCAGGCATCCCAGGCCAGC
CTCTGCCTTCTCCAACTCGGCTTCACCTGGGGGGTGGGCGGAACTCAAATGCCAATGGTTACGTGCGCTTACAAC
TAGGAGGGGAGGACCGGGGAGGGCTCGGGCACCCCCTGCCTGAGCTCGCGGATGAACTGAGACGCAAACTGCAGC
AACGCCAGCCACTGCCCGACTCCAACCCCGAGGAGTCATCAGTATGA**GGGGAACCCCCACCGCGTCGGCGGGAAG**
CGTGGGAG

The MOL5b protein encoded by SEQ ID NO:16 has 833 amino acid residues and is presented using the one-letter code in Table 5D. The Psort profile for MOL5b predicts that this sequence is likely to be localized at the plasma membrane with a certainty of 0.7000.

Table 5D. Encoded MOL5b protein sequence (SEQ ID NO:16)

MAPHWAVWLLAARLWGLGIGAEVWWNLVPRKTVSSGELATVVRRFSQTGIQDFLTLTLTEPTGLLYVGARDHA

```
SALGVPVLLLQAVISWEAPVEKKTECIQKGKNNQTECFNFIRFLQPYNASHLYVCGTYAFQPKCTYVNMLTFT
LEHGEFEDGKGKCPYDPAKGHAGLLVDGELYSATLNNFLGTEPIILRNMGPHHSMKTEYLAFWLNEPHFVGSA
YVPESVGSFTGDDDKVYFFFRERAVESDCYAEQVVARVARVCKGDMGGARTLQRKWTTFLKARLACSAPNWQL
YFNQLQAMHTLQDTSWHNTTFFGVFQAQWGDMYLSAICEYQLEEIQRVFEGPYKEYHEEAQKWDRYTDPVPSP
RPGSCINNWHRRHGYTSSLELPDNILNFVKKHPLMEEQVGPRWSRPLLVKKGTNFTHLVADRVTGLDGATYTV
LFIGTGDGWLLKAVSLGPWVHLIEELQLFDQEPMRSLVLSQSKKLLFAGSRSQLVQLPVADCMKYRSCADCVL
ARDPYCAWSVNTSRCVAVGGHSGSLLIQHVMTSDTSGICNLRGSKKVRPTPKNITVVAGTDLVLPCHLSSNLA
HARWTFGGRDLPAEQPGSFLYDARLQALVVMAAQPRHAGAYHCFSEEQGARLAAEGYLVAVVAGPSVTLEARA
PLENLGLVWLAVVALGAVCLVLLLLVLSLRRRLREELEKGAKATERTLVYPLELPKEPTSPPFRPCPEPDEKL
WDPVGYYYSDGSLKIVPGHARCQPGGGPPSPPPGIPGQPLPSPTRLHLGGGRNSNANGYVRLQLGGEDRGGLG
HPLPELADELRRKLQQRQPLPDSNPEESSV
```

The disclosed nucleic acid sequence for MOL5b has 1695 of 2019 bases (83%) identical to a mouse Semaphorin4C mRNA (GENBANK-ID: S79463) (E= 0.0).

The full MOL5b amino acid sequence has 722 of 834 amino acid residues (86%) identical to, and 765 of 834 residues (91%) positive with the amino acid Semaphorin4C HOMOLOG protein from *Mouse* (S79463_ SEMA_4C_MOUSE) (E= 0.0). The global sequence homology (as defined by FASTA alignment with the full length sequence of this protein) is 91 % amino acid homology and 86 % amino acid identity..

Chromosomal Localization

MOL5b has been localized to human chromosome 2.

MOL5c

In the present invention, the target sequence identified previously, MOL5b, was subjected to the exon linking process to confirm the sequence. PCR primers were designed by starting at the most upstream sequence available, for the forward primer, and at the most downstream sequence available for the reverse primer. In each case, the sequence was examined, walking inward from the respective termini toward the coding sequence, until a suitable sequence that is either unique or highly selective was encountered, or, in the case of the reverse primer, until the stop codon was reached. Such primers were designed based on *in silico* predictions for the full length cDNA, part (one or more exons) of the DNA or protein sequence of the target sequence, or by translated homology of the predicted exons to closely related human sequences sequences from other species. These primers were then employed in PCR amplification based on the following pool of human cDNAs: adrenal gland, bone marrow, brain - amygdala, brain - cerebellum, brain - hippocampus, brain - substantia nigra, brain - thalamus, brain -whole, fetal brain, fetal kidney, fetal liver, fetal lung, heart, kidney, lymphoma - Raji, mammary gland, pancreas, pituitary gland, placenta, prostate, salivary gland, skeletal muscle, small intestine, spinal cord, spleen, stomach, testis, thyroid, trachea, uterus. Usually the resulting amplicons were gel purified, cloned and sequenced to high redundancy. The resulting sequences from all clones were assembled with themselves, with other fragments in CuraGen Corporation's database and with public ESTs. Fragments and

ESTs were included as components for an assembly when the extent of their identity with another component of the assembly was at least 95% over 50 bp. In addition, sequence traces were evaluated manually and edited for corrections if appropriate. These procedures provide the sequence reported below, which is designated MOL5c (Accession Number CG50907-02). This differs from the previously identified sequence, MOL5b, in having 17 different amino acids.

The disclosed novel semaphorin 4C -like nucleic acid of 3112 nucleotides, MOL5c, (also referred to as CG50907-02) is shown in Table 5E. An ORF begins with an ATG initiation codon at nucleotides 104 - 106 and ends with a TGA codon at nucleotides 2603-2605. A putative untranslated region upstream from the initiation codon and downstream from the termination codon is underlined in Table 5E, and the start and stop codons are in bold letters.

Table 5E. MOL5c Nucleotide Sequence (SEQ ID NO:17)

TGCTGCGGGCCCCTCTGGTTTGCTTTCTCTGGCTGTGATTTCTGACCATGTCTTTTCCCTCAGCAGGACAGCTGG
CCTGAAGCTCAGAGCCGGGGCGTGCGCCATGGCCCCACACTGGGCTGTCTGGCTGCTGGCAGCAAGGCTGTGGGG
CCTGGGCATTGGGGCTGAGGTGTGGTGGAACCTTGTGCCGCGTAAGACAGTGTCTTCTGGGGAGCTGGCCACGGT
AGTACGGCGGTTCTCCCAGACCGGCATCCAGGACTTCCTGACACTGACGCTGACGGAGCCCACTGGGCTTCTGTA
CGTGGGCGCCAGGGACCATGCCTCTGCACTGGGCGTCCCTGTGTTGCTGCTGCAGGCTGTGATCTCCTGGGAGGC
CCCCGTGGAGAAGAAGACTGAGTGTATCCAGAAAGGGAAGAACAACCAGACCGAGTGCTTCAACTTCATCCGCTT
CCTGCAGCCCTACAATGCCTCCCACCTGTACGTCTGTGGCACCTACGCCTTCCAGCCCAAGTGCACCTACGTCAA
CATGCTCACCTTCACTTTGGAGCATGGAGAGTTTGAAGATGGGAAGGGCAAGTGTCCCTATGACCCAGCTAAGGG
CCATGCTGGCCTTCTTGTGGATGGTGAGCTGTACTCGGCCACACTCAACAACTTCCTGGGCACGGAACCCATTAT
CCTGCGTAACATGGGGCCCCACCACTCCATGAAGACAGAGTACCTGGCCTTTTGGCTCAACGAACCTCACTTTGT
AGGCTCTGCCTATGTACCTGAGAGTGTGGGCAGCTTCACGGGGGACGACGACAAGGTCTACTTCTTCTTCAGGGA
GCGGGCAGTGGAGTCCGACTGCTATGCCGAGCAGGTGGTGGCTCGTGTGGCCCGTGTCTGCAAGGGCGATATGGG
GGGCGCACGGACCCTGCAGAGGAAGTGGACCACGTTCCTGAAGGCGCGGCTGGCATGCTCTGCCCCGAACTGGCA
GCTCTACTTCAACCAGCTGCAGGCGATGCACACCCTGCAGGACACCTCCTGGCACAACACCACCTTCTTTGGGGT
TTTTCAAGCACAGTGGGGTGACATGTACCTGTCGGCCATCTGTGAGTACCAGTTGGAAGAGATCCAGCGGGTGTT
TGAGGGCCCCTATAAGGAGTACCATGAGGAAGCCCAGAAGTGGGACCGCTACACTGACCCTGTACCCAGCCCTCG
GCCTGGCTCGTGCATTAACAACTGGCATCGGCGCCACGGCTACACCAGCTCCCTGGAGCTACCCGACAACATCCT
CAACTTCGTCAAGAAGCACCCGCTGATGGAGGAGCAGGTGGGGCCTCGGTGGAGCCGCCCCCTGCTCGTGAAGAA
GGGCACCAACTTCACCCACCTGGTGGCCGACCGGGTTACAGGACTTGATGGAGCCACCTATACAGTGCTGTTCAT
TGGCACAGGAGACGGCTGGCTGCTCAAGGCTGTGAGCCTGGGGCCCTGGGTTCACCTGATTGAGGAGCTGCAGCT
GTTTGACCAGGAGCCCATGAGAAGCCTGGTGCTATCTCAGAGCAAGAAGCTGCTCTTTGCCGGCTCCCGCTCTCA
GCTGGTGCAGCTGCCCGTGGCCGACTGCATGAAGTATCGCTCCTGTGCAGACTGTGTCCTCGCCCGGGACCCCTA
TTGCGCCTGGAGCGTCAACACCAGCCGCTGTGTGGCCGTGGGTGGCCACTCTGGATCTCTACTGATCCAGCATGT
GATGACCTCGGACACTTCAGGCATCTGCAACCTCCGTGGCAGTAAGAAAGTCAGGCCCACTCCCAAAAACATCAC
GGTGGTGGCGGGCACAGACCTGGTGCTGCCCTGCCACCTCTCCTCCAACTTGGCCCATGCCCGCTGGACCTTTGG
GGGCCGGGACCTGCCTGCGGAACAGCCCGGGTCCTTCCTCTACGATGCCCGGCTCCAGGCCCTGGTTGTGATGGC
TGCCCAGCCCCGCCATGCCGGGGCCTACCACTGCTTTTCAGAGGAGCAGGGGGCGCGGCTGGCTGCTGAAGGCTA
CCTTGTGGCTGTCGTGGCAGGCCCGTCGGTGACCTTGGAGGCCCGGGCCCCCCTGGAAAACCTGGGGCTGGTGTG
GCTGGCGGTGGTGGCCCTGGGGGCTGTGTGCCTGGTGCTGCTGCTGCTGGTGCTGTCATTGCGCCGGCGGCTGCG
GGAAGAGCTGGAGAAAGGGGCCAAGGCTACTGAGAGGACCTTGGTGTACCCCCTGGAGCTGCCCAAGGAGCCCAC
CAGTCCCCCCTTCCGGCCCTGTCCTGAACCAGATGAGAAACTTTGGGATCCTGTCGGTTACTACTATTCAGATGG
CTCCCTTAAGATAGTACCTGGGCATGCCCGGTGCCAGCCCGGTGGGGGGCCCCCTTCGCCACCTCCAGGCATCCC
AGGCCAGCCTCTGCCTTCTCCAACTCGGCTTCACCTGGGGGGTGGGCGGAACTCAAATGCCAATGGTTACGTGCG
CTTACAACTAGGAGGGGAGGACCGGGGAGGGCTCGGGCACCCCCTGCCTGAGCTCGCGGATGAACTGAGACGCAA
ACTGCAGCAACGCCAGCCACTGCCCGACTCCAACCCCGAGGAGTCATCAGTATGAGGGGAACCCCCACCGCGTCG
GCGGGAAGCGTGGGAGGTGTAGCTCCTACTTTTGCACAGGCACCAGCTACCTCAGGGACATGGCACGGGCACCTG
CTCTGTCTGGGACAGATACTGCCCAGCACCCACCCGGCCATGAGGACCTGCTCTGCTCAGCACGGGCACTGCCAC
TTGGTGTGGCTCACCAGGGCACCAGCCTCGCAGAAGGCATCTTCCTCCTCTCTGTGAATCACAGACACGCGGGAC
CCCAGCCGCCAAAACTTTTCAAGGCAGAAGTTTCAAGATGTGTGTTTGTCTGTATTTGCACATGTGTTTGTGTGT
GTGTGTATGTGTGTGTGCACGCGCGTGCGCGCTTGTGGCATAGCCTTCCTGTTTCTGTCAAGTCTTCCCTTGGCC

TGGGTCCTCCTGGTGAGTCATTGGAGCTATGAAGGGGAAGGGGTCGTATCACTTTGTCTCTCCTACCCCCACTGC
CCCGAGTGTCGGGCAGCGATGTACATATGGAGGTGGG

The MOL5c protein encoded by SEQ ID NO:17 has 833 amino acid residues and is presented using the one-letter code in Table 5F. The Psort profile for MOL5c predicts that this sequence has a signal peptide and the signal peptide is predicted by SignalP to be cleaved between amino acid 20 and 21: GIG-AE. This sequence is likely to be localized at the mitochondrial inner membrane with a certainty of 0.8000 and the plasma membrane with a certainty of 0.7000.

Table 5F. Encoded MOL5c protein sequence (SEQ ID NO:18)

MAPHWAVWLLAARLWGLGIGAEVWWNLVPRKTVSSGELATVVRRFSQTGIQDFLTLTLTEPTGLLYVGARDHA
SALGVPVLLLQAVISWEAPVEKKTECIQKGKNNQTECFNFIRFLQPYNASHLYVCGTYAFQPKCTYVNMLTFT
LEHGEFEDGKGKCPYDPAKGHAGLLVDGELYSATLNNFLGTEPIILRNMGPHHSMKTEYLAFWLNEPHFVGSA
YVPESVGSFTGDDDKVYFFFRERAVESDCYAEQVVARVARVCKGDMGGARTLQRKWTTFLKARLACSAPNWQL
YFNQLQAMHTLQDTSWHNTTFFGVFQAQWGDMYLSAICEYQLEEIQRVFEGPYKEYHEEAQKWDRYTDPVPSP
RPGSCINNWHRRHGYTSSLELPDNILNFVKKHPLMEEQVGPRWSRPLLVKKGTNFTHLVADRVTGLDGATYTV
LFIGTGDGWLLKAVSLGPWVHLIEELQLFDQEPMRSLVLSQSKKLLFAGSRSQLVQLPVADCMKYRSCADCVL
ARDPYCAWSVNTSRCVAVGGHSGSLLIQHVMTSDTSGICNLRGSKKVRPTPKNITVVAGTDLVLPCHLSSNLA
HARWTFGGRDLPAEQPGSFLYDARLQALVVMAAQPRHAGAYHCFSEEQGARLAAEGYLVAVVAGPSVTLEARA
PLENLGLVWLAVVALGAVCLVLLLLVLSLRRRLREELEKGAKATERTLVYPLELPKEPTSPPFRPCPEPDEKL
WDPVGYYYSDGSLKIVPGHARCQPGGGPPSPPPGIPGQPLPSPTRLHLGGGRNSNANGYVRLQLGGEDRGGLG
HPLPELADELRRKLQQRQPLPDSNPEESSV

The disclosed nucleic acid sequence for MOL5c has 2879 of 2906 bases (99%) identical to a gb:GENBANK-ID:AB051526|acc:AB051526.1 mRNA from *Homo sapiens* (*Homo sapiens* mRNA for KIAA1739 protein, partial cds) (E= 0.0).

The full MOL5 amino acid sequence has 722 of 834 amino acid residues (86%) identical to, and 765 of 834 amino acid residues (91%) similar to, the 834 amino acid residue ptnr:SWISSPROT-ACC:Q64151 protein from *Mus musculus* (Mouse) (SEMAPHORIN 4C PRECURSOR (SEMAPHORIN I) (SEMA I) (SEMAPHORIN C-LIKE 1) (M-SEMA F)) (E= 0.0). The global sequence homology (as defined by FASTA alignment with the full length sequence of this protein) is 91% amino acid homology and 86% amino acid identity.

The presence of identifiable domains in the protein disclosed herein was determined by searches versus domain databases such as Pfam, PROSITE, ProDom, Blocks or Prints and then identified by the Interpro domain accession number. Significant domains are summarized in Table 5G.

Table 5G: Domain similarities for MOL5c

Scores for sequence family classification (score includes all domains):

Model	Description	Score	E-value	N
---	---	---	---	---

```
Sema            Sema domain                              664.4   5.8e-196   1
Plexin_repeat   Plexin repeat                             25.8     0.001    1
ig              Immunoglobulin domain                     8.5      0.44    1
integrin_B      Integrins, beta chain                     7.0      0.04    1

Parsed for domains:
Model          Domain  seq-f seq-t    hmm-f hmm-t       score  E-value
--------       ------- ----- -----    ----- -----       -----  -------
Sema             1/1      53   481 ..     1   490 []    664.4 5.8e-196
integrin_B       1/1     505   519 ..     1    14 [.      7.0     0.04
Plexin_repeat    1/1     499   551 ..     1    67 []     25.8    0.001
ig               1/1     570   629 ..     1    45 []      8.5     0.44
```

The Sema domain occurs in semaphorins, which are a large family of secreted and transmembrane proteins, some of which function as repellent signals during axon guidance. Sema domains also occur in a hepatocyte growth factor receptor, in SEX protein (Goodman et al., 1998, Cell 95: 903-916) and in viral proteins.

The presence of these domains indicates that MOL5c likely has properties similar to those of other proteins known to contain this/these domain(s) and similar to the properties of these domains.

Chromosomal Localization

MOL5c maps to chromosome 2. This assignment was made using mapping information associated with genomic clones, public genes and ESTs sharing sequence identity with the disclosed sequence and CuraGen Corporation's Electronic Northern bioinformatic tool.

Tissue Expression

MOL5c is expressed in at least the following tissues: adrenal gland, bone marrow, brain - amygdala, brain - cerebellum, brain - hippocampus, brain - substantia nigra, brain - thalamus, brain -whole, fetal brain, fetal kidney, fetal liver, fetal lung, heart, kidney, lymphoma - Raji, mammary gland, pancreas, pituitary gland, placenta, prostate, salivary gland, skeletal muscle, small intestine, spinal cord, spleen, stomach, testis, thyroid, trachea and uterus. Expression information was derived from the tissue sources of the sequences that were included in the derivation of the sequence of MOL5c.

MOL5a also has homology to other proteins as shown in BLAST alignment results in Table 5H.

Table 5H. BLAST results for MOL5a

Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi|12698023|dbj|BAB 21830.1| (AB051526)	KIAA1739 protein [Homo sapiens]	963	785/801 (98%)	789/801 (98%)	0.0

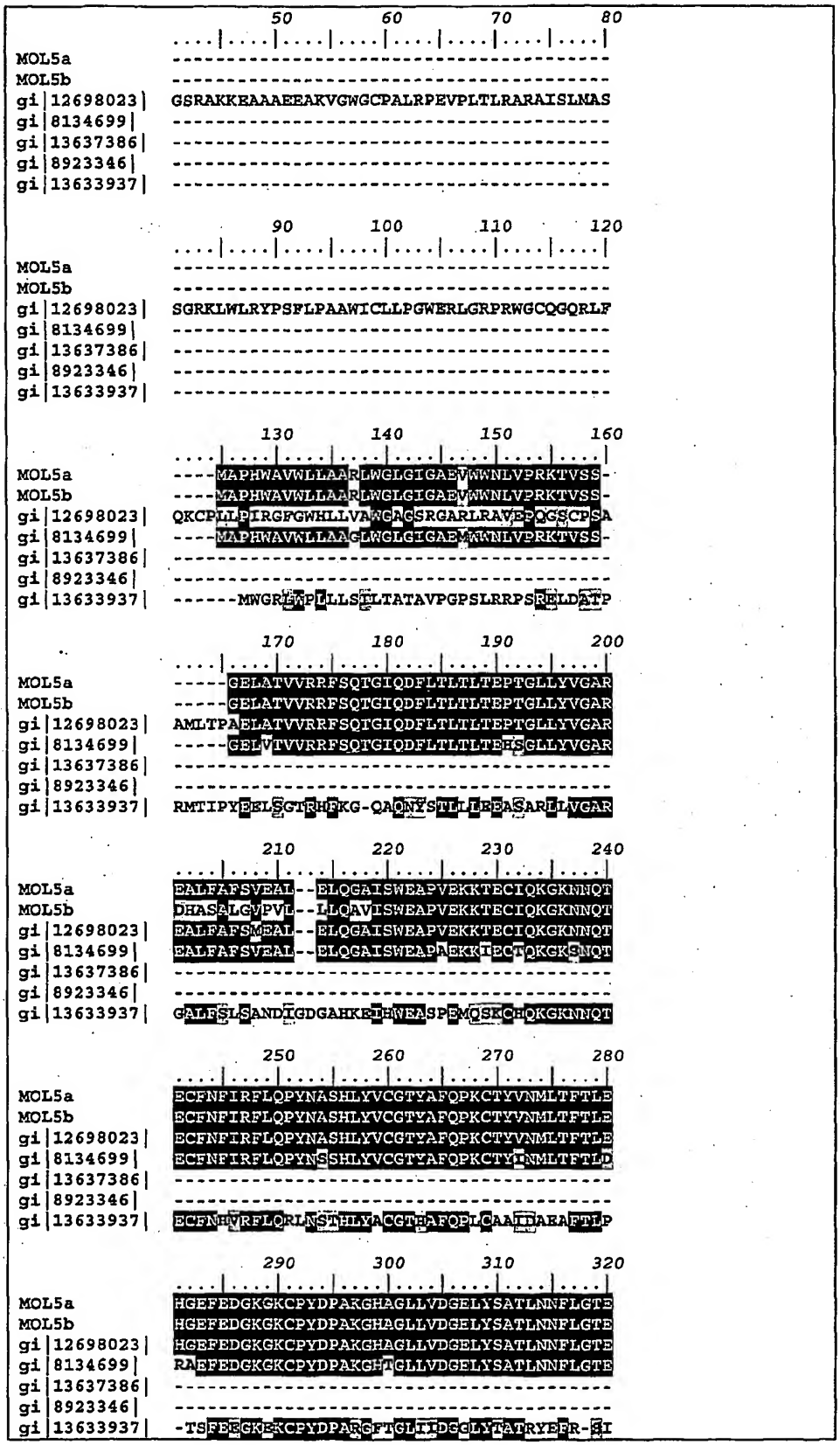
gi|8134699|sp|Q6415 1|SM4C_MOUSE	SEMAPHORIN 4C PRECURSOR (SEMAPHORIN I) (SEMA I) (SEMAPHORIN C-LIKE 1) (M-SEMA F)	834	722/834 (86%)	765/834 (91%)	0.0
gi|13637386|ref|XP_ 002614.2|	hypothetical protein FLJ20369 [Homo sapiens]	510	510/510 (100%)	510/510 (100%)	0.0
gi|8923346|ref|NP_0 60259.1|	sema domain, immunoglobulin domain (Ig), transmembrane domain TM; cytokeratin 14; adipocyte-derived leucine aminopeptidase; hypothetical protein MGC10851; hypothetical protein FLJ14662; sphingomyelin phosphodiesterase -1, acid lysosomal; Pro-platelet ba>	510	509/510 (99%)	509/510 (99%)	0.0
gi|13633937|sp|Q9NT N9|SM4G_HUMAN	SEMAPHORIN 4G PRECURSOR	838	292/673 (43%)	381/673 (56%)	e-138

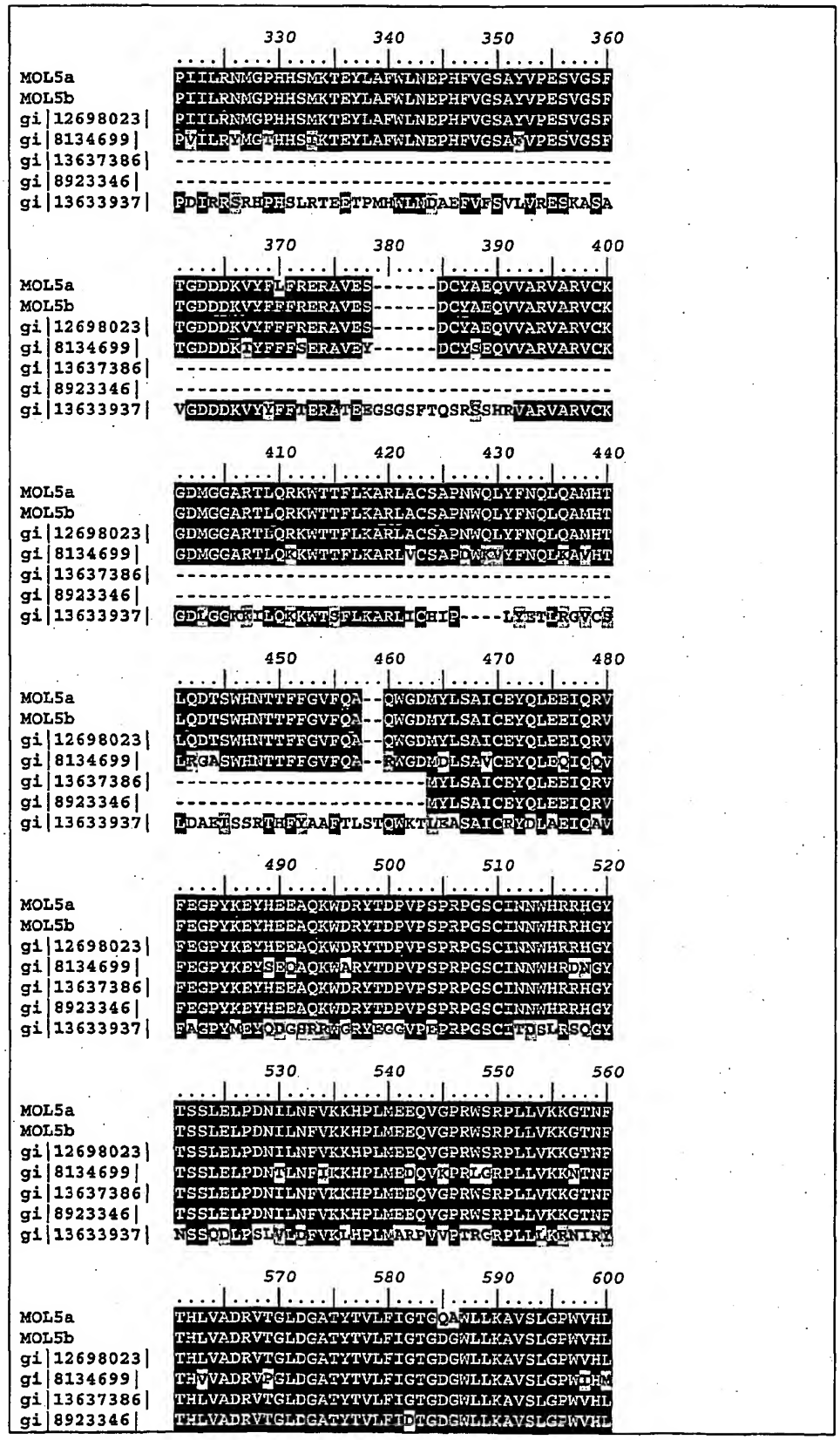
This information is presented graphically in the multiple sequence alignment given in Table 5I (with MOL5a being shown on line 1, and MOL5b on line 2) as a ClustalW analysis comparing MOL5 with related protein sequences.

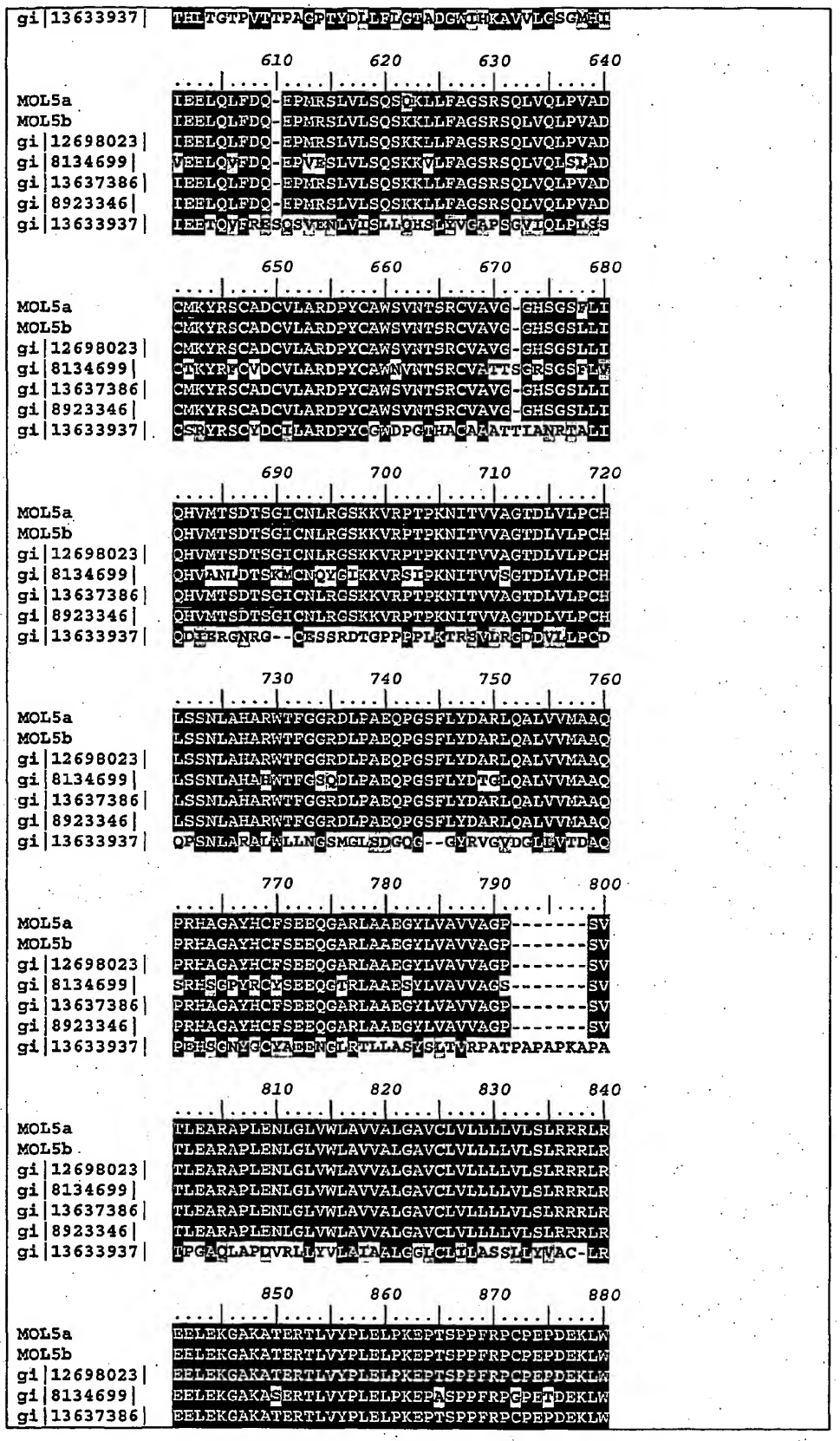
Table 5I Information for the ClustalW proteins:

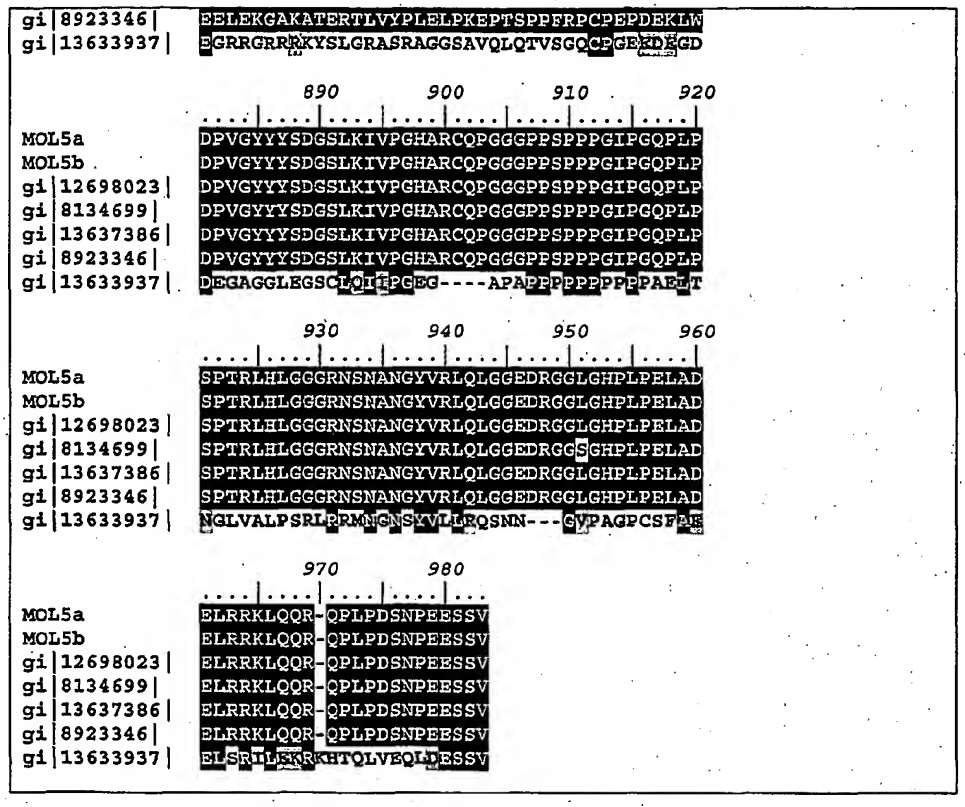
1) MOL5a (SEQ ID NO:14)
2) MOL5b (SEQ ID NO:16)
3) gi|12698023|dbj|BAB21830.1| (AB051526) KIAA1739 protein [Homo sapiens] (SEQ ID NO:46)
4) gi|8134699|sp|Q64151|SM4C_MOUSE SEMAPHORIN 4C PRECURSOR (SEMAPHORIN I) (SEMA I) (SEMAPHORIN C-LIKE 1) (M-SEMA F) (SEQ ID NO:47)
5) gi|8923346|ref|NP_060259.1| sema domain, immunoglobulin domain (Ig), transmembrane domain TM; cytokeratin 14; adipocyte-derived leucine aminopeptidase; hypothetical protein MGC10851; hypothetical protein FLJ14662; sphingomyelin phosphodiesterase-1, acid lysosomal; Pro-platelet ba> (SEQ ID NO:48)
6) gi|13633937|sp|Q9NTN9|SM4G_HUMAN SEMAPHORIN 4G PRECURSOR (SEQ ID NO:49)

```
                        10        20        30        40
                ....|....|....|....|....|....|....|....|
MOL5a           ----------------------------------------
MOL5b           ----------------------------------------
gi|12698023|    LCSHLWQPGLGSCWSEGFPEAGSTHSRLCLLLCWTLIEAV
gi|8134699|     ----------------------------------------
gi|13637386|    ----------------------------------------
gi|8923346|     ----------------------------------------
gi|13633937|    ----------------------------------------
```


50 60 70 80
MOL5a
MOL5b
gi|12698023| GSRAKKEAAAEEAKVGWGCPALRPEVPLTLRARAISLMAS
gi|8134699|
gi|13637386|
gi|8923346|
gi|13633937|
90 100 110 120
MOL5a
MOL5b
gi|12698023| SGRKLWLRYPSFLPAAWICLLPGWERLGRPRWGCQGQRLF
gi|8134699|
gi|13637386|
gi|8923346|
gi|13633937|
130 140 150 160
MOL5a ----MAPHWAVWLLAARLWGLGIGAEVWWNLVPRKTVSS-
MOL5b ----MAPHWAVWLLAARLWGLGIGAEVWWNLVPRKTVSS-
gi|12698023| QKCPLLPIRGFGWHLLVAWCAGSRGARLRAVEPQGSCPSA
gi|8134699| ----MAPHWAVWLLAAGLWGLGIGAEMWWNLVPRKTVSS-
gi|13637386|
gi|8923346|
gi|13633937| ------MWGRLWPLLLSTLTATAVPGPSLRRPSRELDATP
170 180 190 200
MOL5a -----GELATVVRRFSQTGIQDFLTLTLTEPTGLLYVGAR
MOL5b -----GELATVVRRFSQTGIQDFLTLTLTEPTGLLYVGAR
gi|12698023| AMLTPAELATVVRRFSQTGIQDFLTLTLTEPTGLLYVGAR
gi|8134699| -----GELVTVVRRFSQTGIQDFLTLTLTEHSGLLYVGAR
gi|13637386|
gi|8923346|
gi|13633937| RMTIPYEELSGTRHFKG-QAQNYSTLLIEEASARLLVGAR
210 220 230 240
MOL5a EALFAFSVEAL--ELQGAISWEAPVEKKTECIQKGKNNQT
MOL5b DHASALGVPVL--LLQAVISWEAPVEKKTECIQKGKNNQT
gi|12698023| EALFAFSMEAL--ELQGAISWEAPVEKKTECIQKGKNNQT
gi|8134699| EALFAFSVEAL--ELQGAISWEAPAEKKIECTQKGKSNQT
gi|13637386|
gi|8923346|
gi|13633937| GALFSLSANDIGDGAHKELHWEASPEMQSKCHQKGKNNQT
250 260 270 280
MOL5a ECFNFIRFLQPYNASHLYVCGTYAFQPKCTYVNMLTFTLE
MOL5b ECFNFIRFLQPYNASHLYVCGTYAFQPKCTYVNMLTFTLE
gi|12698023| ECFNFIRFLQPYNASHLYVCGTYAFQPKCTYVNMLTFTLE
gi|8134699| ECFNFIRFLQPYNSSHLYVCGTYAFQPKCTYTNMLTFTLD
gi|13637386|
gi|8923346|
gi|13633937| ECFNHVRFLQRLNSTHLYACGTEAFQPLCAAIDAEAFTLP
290 300 310 320
MOL5a HGEFEDGKGKCPYDPAKGHAGLLVDGELYSATLNNFLGTE
MOL5b HGEFEDGKGKCPYDPAKGHAGLLVDGELYSATLNNFLGTE
gi|12698023| HGEFEDGKGKCPYDPAKGHAGLLVDGELYSATLNNFLGTE
gi|8134699| RAEFEDGKGKCPYDPAKGHTGLLVDGELYSATLNNFLGTE
gi|13637386|
gi|8923346|
gi|13633937| -TSFEEGKEKCPYDPARGFTGLIIDGGLYTATRYEFR-SI

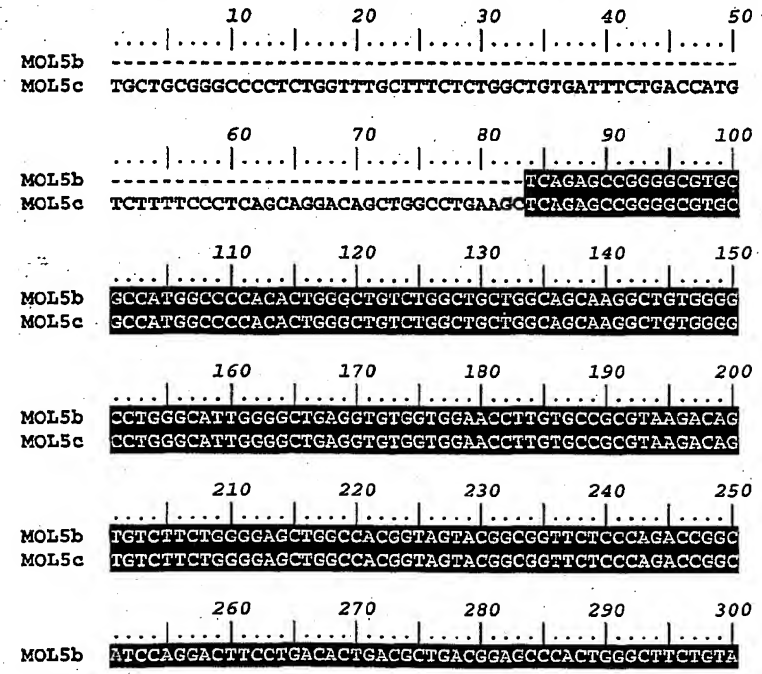

330 340 350 360
MOL5a PIILRNMGPHHSMKTEYLAFWLNEPHFVGSAYVPESVGSF
MOL5b PIILRNMGPHHSMKTEYLAFWLNEPHFVGSAYVPESVGSF
gi|12698023| PIILRNMGPHHSMKTEYLAFWLNEPHFVGSAYVPESVGSF
gi|8134699| PVILRYMGTHHSMKTEYLAFWLNEPHFVGSAGVPESVGSF
gi|13637386| ----------
gi|8923346| ----------
gi|13633937| PDIRRSRHPHSLRTEETPMHNLNDAEFVFSVLVRESKASA
370 380 390 400
MOL5a TGDDDKVYFLFRERAVES------DCYAEQVVARVARVCK
MOL5b TGDDDKVYFFFRERAVES------DCYAEQVVARVARVCK
gi|12698023| TGDDDKVYFFFRERAVES------DCYAEQVVARVARVCK
gi|8134699| TGDDDKIYFFFSERAVEY------DCYSEQVVARVARVCK
gi|13637386| ----------
gi|8923346| ----------
gi|13633937| VGDDDKVYYFFTERATEEGSGSFTQSRSSHRVARVARVCK
410 420 430 440
MOL5a GDMGGARTLQRKWTTFLKARLACSAPNWQLYFNQLQAMHT
MOL5b GDMGGARTLQRKWTTFLKARLACSAPNWQLYFNQLQAMHT
gi|12698023| GDMGGARTLQRKWTTFLKARLACSAPNWQLYFNQLQAMHT
gi|8134699| GDMGGARTLQKKWTTFLKARLVCSAPDWKVYFNQLKAVHT
gi|13637386| ----------
gi|8923346| ----------
gi|13633937| GDVGGGKILQKKWTSFLKARLICHIP----LYETLRGVCS
450 460 470 480
MOL5a LQDTSWHNTTFFGVFQA--QWGDMYLSAICEYQLEEIQRV
MOL5b LQDTSWHNTTFFGVFQA--QWGDMYLSAICEYQLEEIQRV
gi|12698023| LQDTSWHNTTFFGVFQA--QWGDMYLSAICEYQLEEIQRV
gi|8134699| LRGASWHNTTFFGVFQA--RWGDMDLSAVCEYQLEQIQQV
gi|13637386| ----------MYLSAICEYQLEEIQRV
gi|8923346| ----------MYLSAICEYQLEEIQRV
gi|13633937| LDAETSSRTHFYAAFTLSTQWKTLEASAICRYDLAEIQAV
490 500 510 520
MOL5a FEGPYKEYHEEAQKWDRYTDPVPSPRPGSCINNWHRRHGY
MOL5b FEGPYKEYHEEAQKWDRYTDPVPSPRPGSCINNWHRRHGY
gi|12698023| FEGPYKEYHEEAQKWDRYTDPVPSPRPGSCINNWHRRHGY
gi|8134699| FEGPYKEYSEQAQKWARYTDPVPSPRPGSCINNWHRDNGY
gi|13637386| FEGPYKEYHEEAQKWDRYTDPVPSPRPGSCINNWHRRHGY
gi|8923346| FEGPYKEYHEEAQKWDRYTDPVPSPRPGSCINNWHRRHGY
gi|13633937| FAGPYMEYQDGSRRWGRYEGGVPEPRPGSCITDSLRSQGY
530 540 550 560
MOL5a TSSLELPDNILNFVKKHPLMEEQVGPRWSRPLLVKKGTNF
MOL5b TSSLELPDNILNFVKKHPLMEEQVGPRWSRPLLVKKGTNF
gi|12698023| TSSLELPDNILNFVKKHPLMEEQVGPRWSRPLLVKKGTNF
gi|8134699| TSSLELPDNTLNFIKKHPLMEDQVKPRLGRPLLVKKNTNF
gi|13637386| TSSLELPDNILNFVKKHPLMEEQVGPRNSRPLLVKKGTNF
gi|8923346| TSSLELPDNILNFVKKHPLMEEQVGPRWSRPLLVKKGTNF
gi|13633937| NSSQDLPSLVLDFVKLHPLMARPVVPTRGRPLLVKRNIRY
570 580 590 600
MOL5a THLVADRVTGLDGATYTVLFIGTGQAWLLKAVSLGPWVHL
MOL5b THLVADRVTGLDGATYTVLFIGTGDGWLLKAVSLGPWVHL
gi|12698023| THLVADRVTGLDGATYTVLFIGTGDGWLLKAVSLGPWVHL
gi|8134699| THVVADRVPGLDGATYTVLFIGTGDGWLLKAVSLGPWQHM
gi|13637386| THLVADRVTGLDGATYTVLFIGTGDGWLLKAVSLGPWVHL
gi|8923346| THLVADRVTGLDGATYTVLFIDTGDGWLLKAVSLGPWVHL

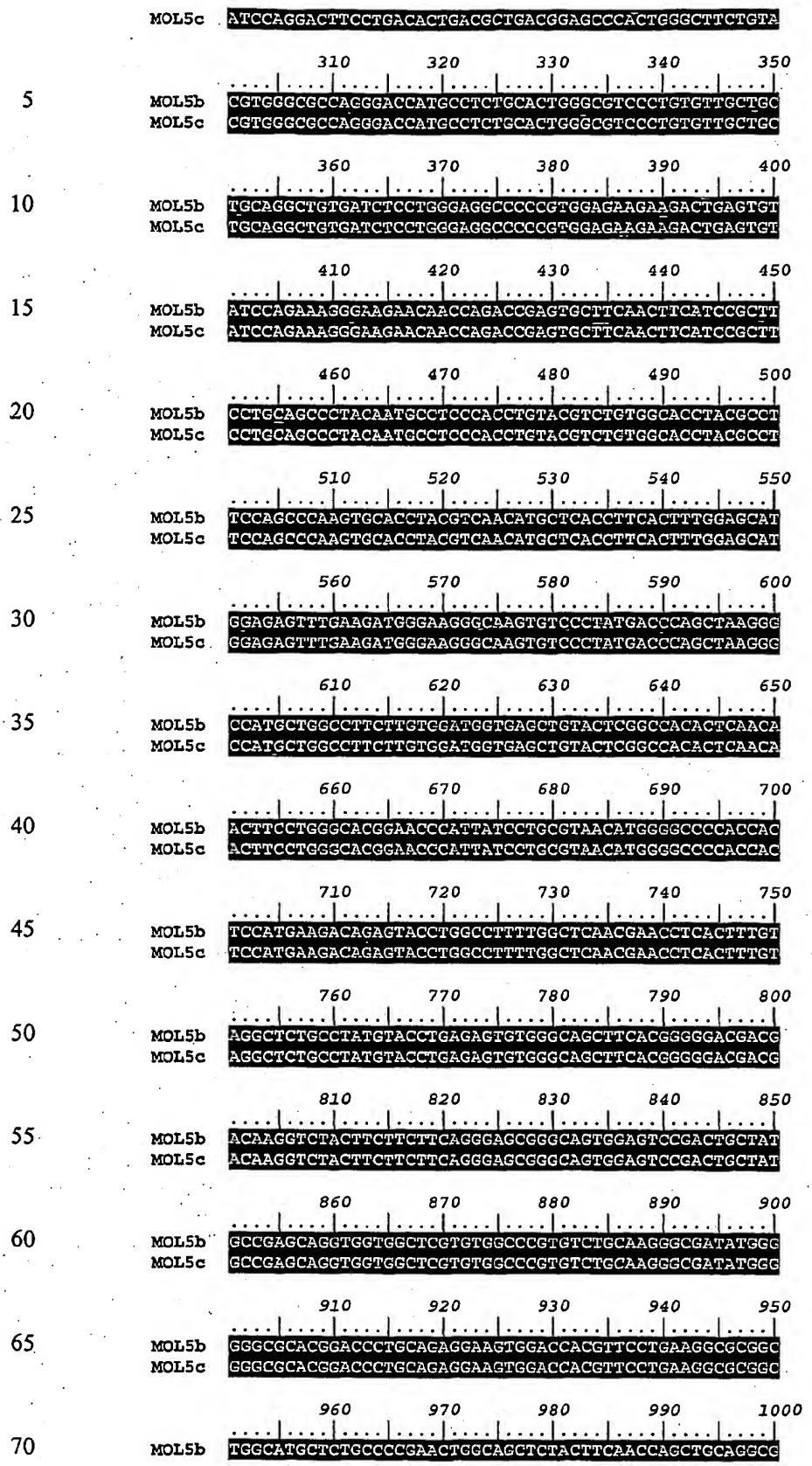

gi|13633937|
610 620 630 640
MOL5a
MOL5b
gi|12698023|
gi|8134699|
gi|13637386|
gi|8923346|
gi|13633937|
650 660 670 680
MOL5a
MOL5b
gi|12698023|
gi|8134699|
gi|13637386|
gi|8923346|
gi|13633937|
690 700 710 720
MOL5a
MOL5b
gi|12698023|
gi|8134699|
gi|13637386|
gi|8923346|
gi|13633937|
730 740 750 760
MOL5a
MOL5b
gi|12698023|
gi|8134699|
gi|13637386|
gi|8923346|
gi|13633937|
770 780 790 800
MOL5a
MOL5b
gi|12698023|
gi|8134699|
gi|13637386|
gi|8923346|
gi|13633937|
810 820 830 840
MOL5a
MOL5b
gi|12698023|
gi|8134699|
gi|13637386|
gi|8923346|
gi|13633937|
850 860 870 880
MOL5a
MOL5b
gi|12698023|
gi|8134699|
gi|13637386|



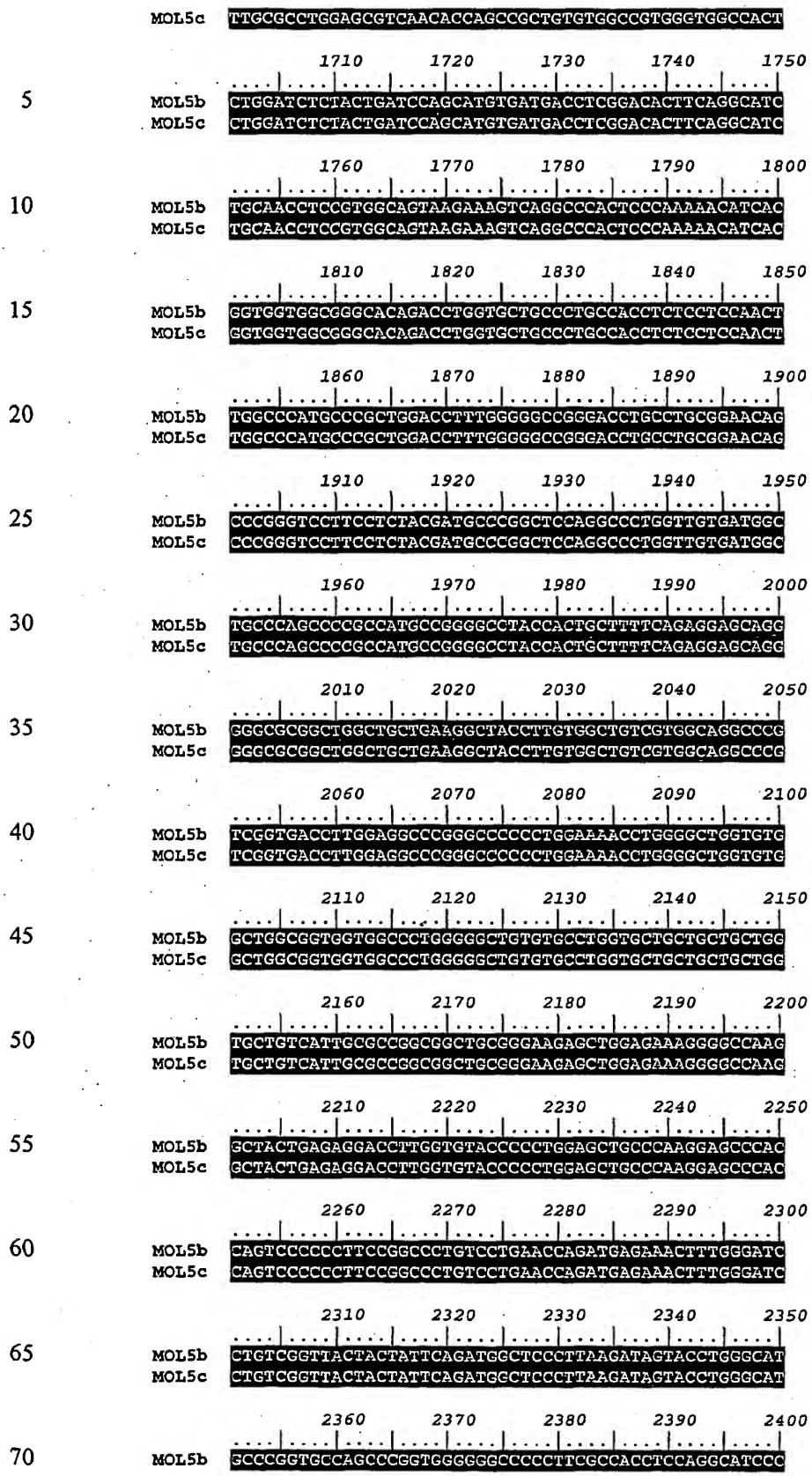
MOL5b and MOL5c share close homology to each other and therefore to other proteins as is shown in the BLAST alignment in Table 5J.

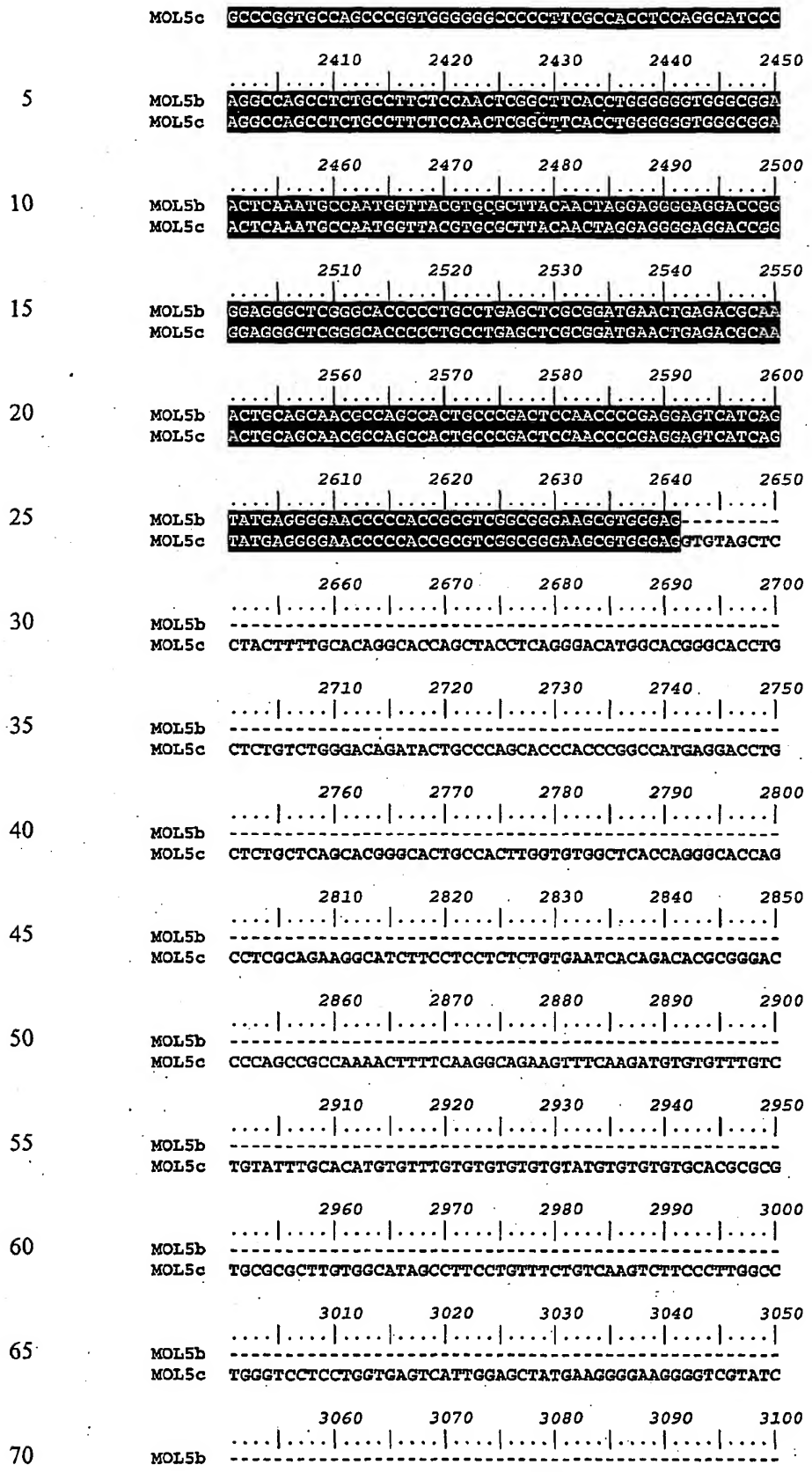
Table 5J. BLAST alignment between MOL5b and MOL5c




MOL5c ATCCAGGACTTCCTGACACTGACGCTGACGGAGCCCACTGGGCTTCTGTA
310 320 330 340 350
MOL5b CGTGGGCGCCAGGGACCATGCCTCTGCACTGGGCGTCCCTGTGTTGCTGC
MOL5c CGTGGGCGCCAGGGACCATGCCTCTGCACTGGGCGTCCCTGTGTTGCTGC
360 370 380 390 400
MOL5b TGCAGGCTGTGATCTCCTGGGAGGCCCCCGTGGAGAAGAAGACTGAGTGT
MOL5c TGCAGGCTGTGATCTCCTGGGAGGCCCCCGTGGAGAAGAAGACTGAGTGT
410 420 430 440 450
MOL5b ATCCAGAAAGGGAAGAACAACCAGACCGAGTGCTTCAACTTCATCCGCTT
MOL5c ATCCAGAAAGGGAAGAACAACCAGACCGAGTGCTTCAACTTCATCCGCTT
460 470 480 490 500
MOL5b CCTGCAGCCCTACAATGCCTCCCACCTGTACGTCTGTGGCACCTACGCCT
MOL5c CCTGCAGCCCTACAATGCCTCCCACCTGTACGTCTGTGGCACCTACGCCT
510 520 530 540 550
MOL5b TCCAGCCCAAGTGCACCTACGTCAACATGCTCACCTTCACTTTGGAGCAT
MOL5c TCCAGCCCAAGTGCACCTACGTCAACATGCTCACCTTCACTTTGGAGCAT
560 570 580 590 600
MOL5b GGAGAGTTTGAAGATGGGAAGGGCAAGTGTCCCTATGACCCAGCTAAGGG
MOL5c GGAGAGTTTGAAGATGGGAAGGGCAAGTGTCCCTATGACCCAGCTAAGGG
610 620 630 640 650
MOL5b CCATGCTGGCCTTCTTGTGGATGGTGAGCTGTACTCGGCCACACTCAACA
MOL5c CCATGCTGGCCTTCTTGTGGATGGTGAGCTGTACTCGGCCACACTCAACA
660 670 680 690 700
MOL5b ACTTCCTGGGCACGGAACCCATTATCCTGCGTAACATGGGGCCCCACCAC
MOL5c ACTTCCTGGGCACGGAACCCATTATCCTGCGTAACATGGGGCCCCACCAC
710 720 730 740 750
MOL5b TCCATGAAGACAGAGTACCTGGCCTTTTGGCTCAACGAACCTCACTTTGT
MOL5c TCCATGAAGACAGAGTACCTGGCCTTTTGGCTCAACGAACCTCACTTTGT
760 770 780 790 800
MOL5b AGGCTCTGCCTATGTACCTGAGAGTGTGGGCAGCTTCACGGGGGACGACG
MOL5c AGGCTCTGCCTATGTACCTGAGAGTGTGGGCAGCTTCACGGGGGACGACG
810 820 830 840 850
MOL5b ACAAGGTCTACTTCTTCTTCAGGGAGCGGGCAGTGGAGTCCGACTGCTAT
MOL5c ACAAGGTCTACTTCTTCTTCAGGGAGCGGGCAGTGGAGTCCGACTGCTAT
860 870 880 890 900
MOL5b GCCGAGCAGGTGGTGGCTCGTGTGGCCCGTGTCTGCAAGGGCGATATGGG
MOL5c GCCGAGCAGGTGGTGGCTCGTGTGGCCCGTGTCTGCAAGGGCGATATGGG
910 920 930 940 950
MOL5b GGGCGCACGGACCCTGCAGAGGAAGTGGACCACGTTCCTGAAGGCGCGGC
MOL5c GGGCGCACGGACCCTGCAGAGGAAGTGGACCACGTTCCTGAAGGCGCGGC
960 970 980 990 1000
MOL5b TGGCATGCTCTGCCCCGAACTGGCAGCTCTACTTCAACCAGCTGCAGGCG

```
MOL5c TGGCATGCTCTGCCCCGAACTGGCAGCTCTACTTCAACCAGCTGCAGGCG

                1010        1020        1030        1040        1050
                ....|....|....|....|....|....|....|....|....|....|
MOL5b ATGCACACCCTGCAGGACACCTCCTGGCACAACACCACCTTCTTTGGGGT
MOL5c ATGCACACCCTGCAGGACACCTCCTGGCACAACACCACCTTCTTTGGGGT

                1060        1070        1080        1090        1100
                ....|....|....|....|....|....|....|....|....|....|
MOL5b TTTTCAAGCACAGTGGGGTGACATGTACCTGTCGGCCATCTGTGAGTACC
MOL5c TTTTCAAGCACAGTGGGGTGACATGTACCTGTCGGCCATCTGTGAGTACC

                1110        1120        1130        1140        1150
                ....|....|....|....|....|....|....|....|....|....|
MOL5b AGTTGGAAGAGATCCAGCGGGTGTTTGAGGGCCCCTATAAGGAGTACCAT
MOL5c AGTTGGAAGAGATCCAGCGGGTGTTTGAGGGCCCCTATAAGGAGTACCAT

                1160        1170        1180        1190        1200
                ....|....|....|....|....|....|....|....|....|....|
MOL5b GAGGAAGCCCAGAAGTGGGACCGCTACACTGACCCTGTACCCAGCCCTCG
MOL5c GAGGAAGCCCAGAAGTGGGACCGCTACACTGACCCTGTACCCAGCCCTCG

                1210        1220        1230        1240        1250
                ....|....|....|....|....|....|....|....|....|....|
MOL5b GCCTGGCTCGTGCATTAACAACTGGCATCGGCGCCACGGCTACACCAGCT
MOL5c GCCTGGCTCGTGCATTAACAACTGGCATCGGCGCCACGGCTACACCAGCT

                1260        1270        1280        1290        1300
                ....|....|....|....|....|....|....|....|....|....|
MOL5b CCCTGGAGCTACCCGACAACATCCTCAACTTCGTCAAGAAGCACCCGCTG
MOL5c CCCTGGAGCTACCCGACAACATCCTCAACTTCGTCAAGAAGCACCCGCTG

                1310        1320        1330        1340        1350
                ....|....|....|....|....|....|....|....|....|....|
MOL5b ATGGAGGAGCAGGTGGGGCCTCGGTGGAGCCGCCCCCTGCTCGTGAAGAA
MOL5c ATGGAGGAGCAGGTGGGGCCTCGGTGGAGCCGCCCCCTGCTCGTGAAGAA

                1360        1370        1380        1390        1400
                ....|....|....|....|....|....|....|....|....|....|
MOL5b GGGCACCAACTTCACCCACCTGGTGGCCGACCGGGTTACAGGACTTGATG
MOL5c GGGCACCAACTTCACCCACCTGGTGGCCGACCGGGTTACAGGACTTGATG

                1410        1420        1430        1440        1450
                ....|....|....|....|....|....|....|....|....|....|
MOL5b GAGCCACCTATACAGTGCTGTTCATTGGCACAGGAGACGGCTGGCTGCTC
MOL5c GAGCCACCTATACAGTGCTGTTCATTGGCACAGGAGACGGCTGGCTGCTC

                1460        1470        1480        1490        1500
                ....|....|....|....|....|....|....|....|....|....|
MOL5b AAGGCTGTGAGCCTGGGGCCCTGGGTTCACCTGATTGAGGAGCTGCAGCT
MOL5c AAGGCTGTGAGCCTGGGGCCCTGGGTTCACCTGATTGAGGAGCTGCAGCT

                1510        1520        1530        1540        1550
                ....|....|....|....|....|....|....|....|....|....|
MOL5b GTTTGACCAGGAGCCCATGAGAAGCCTGGTGCTATCTCAGAGCAAGAAGC
MOL5c GTTTGACCAGGAGCCCATGAGAAGCCTGGTGCTATCTCAGAGCAAGAAGC

                1560        1570        1580        1590        1600
                ....|....|....|....|....|....|....|....|....|....|
MOL5b TGCTCTTTGCCGGCTCCCGCTCTCAGCTGGTGCAGCTGCCCGTGGCCGAC
MOL5c TGCTCTTTGCCGGCTCCCGCTCTCAGCTGGTGCAGCTGCCCGTGGCCGAC

                1610        1620        1630        1640        1650
                ....|....|....|....|....|....|....|....|....|....|
MOL5b TGCATGAAGTATCGCTCCTGTGCAGACTGTGTCCTCGCCCGGGACCCCTA
MOL5c TGCATGAAGTATCGCTCCTGTGCAGACTGTGTCCTCGCCCGGGACCCCTA

                1660        1670        1680        1690        1700
                ....|....|....|....|....|....|....|....|....|....|
MOL5b TTGCGCCTGGAGCGTCAACACCAGCCGCTGTGTGGCCGTGGGTGGCCACT
```


MOL5c TTGCGCCTGGAGCGTCAACACCAGCCGCTGTGTGGCCGTGGGTGGCCACT
1710 1720 1730 1740 1750
MOL5b CTGGATCTCTACTGATCCAGCATGTGATGACCTCGGACACTTCAGGCATC
MOL5c CTGGATCTCTACTGATCCAGCATGTGATGACCTCGGACACTTCAGGCATC
1760 1770 1780 1790 1800
MOL5b TGCAACCTCCGTGGCAGTAAGAAAGTCAGGCCCACTCCCAAAAACATCAC
MOL5c TGCAACCTCCGTGGCAGTAAGAAAGTCAGGCCCACTCCCAAAAACATCAC
1810 1820 1830 1840 1850
MOL5b GGTGGTGGCGGGCACAGACCTGGTGCTGCCCTGCCACCTCTCCTCCAACT
MOL5c GGTGGTGGCGGGCACAGACCTGGTGCTGCCCTGCCACCTCTCCTCCAACT
1860 1870 1880 1890 1900
MOL5b TGGCCCATGCCCGCTGGACCTTTGGGGGCCGGGACCTGCCTGCGGAACAG
MOL5c TGGCCCATGCCCGCTGGACCTTTGGGGGCCGGGACCTGCCTGCGGAACAG
1910 1920 1930 1940 1950
MOL5b CCCGGGTCCTTCCTCTACGATGCCCGGCTCCAGGCCCTGGTTGTGATGGC
MOL5c CCCGGGTCCTTCCTCTACGATGCCCGGCTCCAGGCCCTGGTTGTGATGGC
1960 1970 1980 1990 2000
MOL5b TGCCCAGCCCCGCCATGCCGGGGCCTACCACTGCTTTTCAGAGGAGCAGG
MOL5c TGCCCAGCCCCGCCATGCCGGGGCCTACCACTGCTTTTCAGAGGAGCAGG
2010 2020 2030 2040 2050
MOL5b GGGCGCGGCTGGCTGCTGAAGGCTACCTTGTGGCTGTCGTGGCAGGCCCG
MOL5c GGGCGCGGCTGGCTGCTGAAGGCTACCTTGTGGCTGTCGTGGCAGGCCCG
2060 2070 2080 2090 2100
MOL5b TCGGTGACCTTGGAGGCCCGGGCCCCCCTGGAAAACCTGGGGCTGGTGTG
MOL5c TCGGTGACCTTGGAGGCCCGGGCCCCCCTGGAAAACCTGGGGCTGGTGTG
2110 2120 2130 2140 2150
MOL5b GCTGGCGGTGGTGGCCCTGGGGGCTGTGTGCCTGGTGCTGCTGCTGCTGG
MOL5c GCTGGCGGTGGTGGCCCTGGGGGCTGTGTGCCTGGTGCTGCTGCTGCTGG
2160 2170 2180 2190 2200
MOL5b TGCTGTCATTGCGCCGGCGGCTGCGGGAAGAGCTGGAGAAAGGGGCCAAG
MOL5c TGCTGTCATTGCGCCGGCGGCTGCGGGAAGAGCTGGAGAAAGGGGCCAAG
2210 2220 2230 2240 2250
MOL5b GCTACTGAGAGGACCTTGGTGTACCCCCTGGAGCTGCCCAAGGAGCCCAC
MOL5c GCTACTGAGAGGACCTTGGTGTACCCCCTGGAGCTGCCCAAGGAGCCCAC
2260 2270 2280 2290 2300
MOL5b CAGTCCCCCCTTCCGGCCCTGTCCTGAACCAGATGAGAAACTTTGGGATC
MOL5c CAGTCCCCCCTTCCGGCCCTGTCCTGAACCAGATGAGAAACTTTGGGATC
2310 2320 2330 2340 2350
MOL5b CTGTCGGTTACTACTATTCAGATGGCTCCCTTAAGATAGTACCTGGGCAT
MOL5c CTGTCGGTTACTACTATTCAGATGGCTCCCTTAAGATAGTACCTGGGCAT
2360 2370 2380 2390 2400
MOL5b GCCCGGTGCCAGCCCGGTGGGGGGCCCCCTTCGCCACCTCCAGGCATCCC


MOL5c GCCCGGTGCCAGCCCGGTGGGGGGCCCCCTTCGCCACCTCCAGGCATCCC
2410 2420 2430 2440 2450
MOL5b AGGCCAGCCTCTGCCTTCTCCAACTCGGCTTCACCTGGGGGGTGGGCGGA
MOL5c AGGCCAGCCTCTGCCTTCTCCAACTCGGCTTCACCTGGGGGGTGGGCGGA
2460 2470 2480 2490 2500
MOL5b ACTCAAATGCCAATGGTTACGTGCGCTTACAACTAGGAGGGGAGGACCGG
MOL5c ACTCAAATGCCAATGGTTACGTGCGCTTACAACTAGGAGGGGAGGACCGG
2510 2520 2530 2540 2550
MOL5b GGAGGGCTCGGGCACCCCCTGCCTGAGCTCGCGGATGAACTGAGACGCAA
MOL5c GGAGGGCTCGGGCACCCCCTGCCTGAGCTCGCGGATGAACTGAGACGCAA
2560 2570 2580 2590 2600
MOL5b ACTGCAGCAACCCCAGCCACTGCCCGACTCCAACCCCGAGGAGTCATCAG
MOL5c ACTGCAGCAACGCCAGCCACTGCCCGACTCCAACCCCGAGGAGTCATCAG
2610 2620 2630 2640 2650
MOL5b TATGAGGGGAACCCCCACCGCGTCGGCGGGAAGCGTGGGAG---------
MOL5c TATGAGGGGAACCCCCACCGCGTCGGCGGGAAGCGTGGGAGGTGTAGCTC
2660 2670 2680 2690 2700
MOL5b ----------
MOL5c CTACTTTTGCACAGGCACCAGCTACCTCAGGGACATGGCACGGGCACCTG
2710 2720 2730 2740 2750
MOL5b ----------
MOL5c CTCTGTCTGGGACAGATACTGCCCAGCACCCACCCGGCCATGAGGACCTG
2760 2770 2780 2790 2800
MOL5b ----------
MOL5c CTCTGCTCAGCACGGGCACTGCCACTTGGTGTGGCTCACCAGGGCACCAG
2810 2820 2830 2840 2850
MOL5b ----------
MOL5c CCTCGCAGAAGGCATCTTCCTCCTCTCTGTGAATCACAGACACGCGGGAC
2860 2870 2880 2890 2900
MOL5b ----------
MOL5c CCCAGCCGCCAAAACTTTTCAAGGCAGAAGTTTCAAGATGTGTGTTTGTC
2910 2920 2930 2940 2950
MOL5b ----------
MOL5c TGTATTTGCACATGTGTTTGTGTGTGTGTGTATGTGTGTGTGCACGCGCG
2960 2970 2980 2990 3000
MOL5b ----------
MOL5c TGCGCGCTTGTGGCATAGCCTTCCTGTTTCTGTCAAGTCTTCCCTTGGCC
3010 3020 3030 3040 3050
MOL5b ----------
MOL5c TGGGTCCTCCTGGTGAGTCATTGGAGCTATGAAGGGGAAGGGGTCGTATC
3060 3070 3080 3090 3100
MOL5b ----------

```
MOL5c  ACTTTGTCTCTCCTACCCCCACTGCCCCGAGTGTCGGGCAGCGATGTACA

                3110
           ....|....|..
MOL5b      -----------
MOL5c      TATGGAGGTGGG
```

As used herein, any reference to MOL5 encompasses MOL5a, MOL5b, and MOL5c, unless otherwise indicated.

Table 5K and 5L list the domain descriptions from DOMAIN analysis results against MOL5. The region from amino acid residue 66 through 487 (SEQ ID NO:14) most probably ($E = 3e^{-125}$) contains a Sema domain found in Semaphorins, described above under MOL4, and aligned here in Table 5K. The region from amino acid residue 562 through 627 (SEQ ID NO:14) most probably ($E = 1e^{-4}$) also contains a Sema domain found in Semaphorins, aligned here in Table 5L. This indicates that the MOL5 sequence has properties similar to those of other proteins known to contain this domain.

Table 5K. Domain Analysis of MOL5

gnl|Pfam|pfam01403, Sema, Sema domain.
CD-Length = 431 residues, 97.0% aligned
Score = 442 bits (1137), Expect = 3e-125

```
                      10        20        30        40        50
                ....|....|....|....|....|....|....|....|....|....|
MOL5_1          ELILTLTLSPTGLLYVGARDHASALGVPVL--T-LQAVISNPADVEKKTEC
Pfam|pfam01403  FRTLLDDEDRGRLYVGARNHVYVINIEDLSEV-LNLKIGWPASCEKCEEC

                      60        70        80        90       100
                ....|....|....|....|....|....|....|....|....|....|
MOL5_1          IQKGRNNQTECFNFIRVLQPYNASHLYVCGIYAFQPKCTYVNMLT-----
Pfam|pfam01403  NMKGKSPLTECTNFIRVLQAYNDTHLYVCGTNAFNPKCTLTNLGD-----

                     110       120       130       140       150
                ....|....|....|....|....|....|....|....|....|....|
MOL5_1          ------------------------------FTLEHGEFEDGKGKCPYD
Pfam|pfam01403  ------------------------------LFSLDMDNEESGCGDCPYD

                     160       170       180       190       200
                ....|....|....|....|....|....|....|....|....|....|
MOL5_1          PAKGHAGLLVDGELYSATLNNFLGTEPILLRNMG---PEHSMKTE-YLAF
Pfam|pfam01403  PLGNTTSVLVGGELYSGTAIDFSGRDFSIRRLLG---SEDGARTEFFDSK

                     210       220       230       240       250
                ....|....|....|....|....|....|....|....|....|....|
MOL5_1          WLNEPHFVGSAYVPESVGSFTG---DDKVYFFFRETAVESD-CYAFQVV
Pfam|pfam01403  WLNLPNFVDSYPIHYVH---SF---SDDKVYFFFRETAVEYS-NC-KAIH

                     260       270       280       290       300
                ....|....|....|....|....|....|....|....|....|....|
MOL5_1          ARVARVCKGDVGGARILQRKWTTFLKARLACSAP---NWQLVFNQLQAMH
Pfam|pfam01403  SRVARVCKNDPGGHSYLENKWTTFLKARLNCSIPG-EGTPFYFNGLQAAF

                     310       320       330       340       350
                ....|....|....|....|....|....|....|....|....|....|
MOL5_1          TLQDTEWHNTTFFGVEQAQWGDMYL--SAICEVQLEDLQRVF-EGEVKEY
Pfam|pfam01403  LLPTDNDTDPVLVGVETTSSNSSAG--SAVCAFSMKDINQVF-EGPFKHQ

                     360       370       380       390       400
                ....|....|....|....|....|....|....|....|....|....|
MOL5_1          ----HEEAQKWDRYTDPVPSPRPGSFINNWHRRHGYTSSL--------EL
```

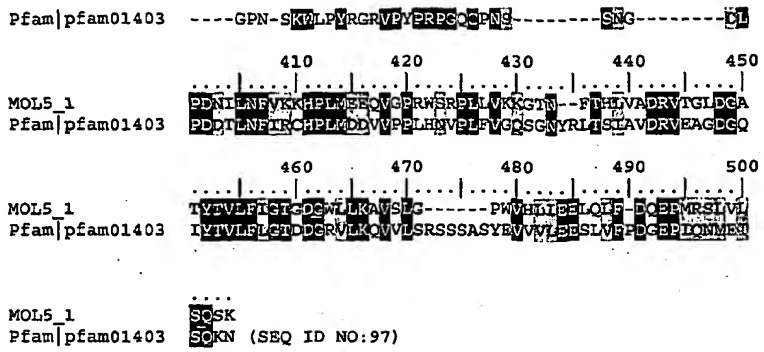
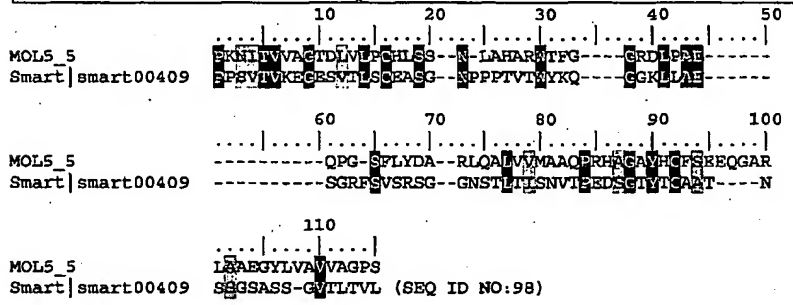


Table 5L. Domain Analysis of MOL5

gnl|Smart|smart00409, IG, Immunoglobulin
CD-Length = 86 residues, only 79.1% aligned
Score = 42.0 bits (97), Expect = 1e-04



The protein similarity information, expression pattern, cellular localization, and map location for MOL5 suggest that this Semaphorin 4C-like protein may have important structural and/or physiological functions characteristic of the Semaphorin family. These functions include growth cone guidance, axonal pathfindin, and embryonic development. Therefore, the MOL5 nucleic acids and proteins are useful in potential diagnostic and therapeutic applications and as a research tool. These include serving as a specific or selective nucleic acid or protein diagnostic and/or prognostic marker, wherein the presence or amount of the nucleic acid or the protein are to be assessed. These also include potential therapeutic applications such as the following: (i) a protein therapeutic, (ii) a small molecule drug target, (iii) an antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), (iv) a nucleic acid useful in gene therapy (gene delivery/gene ablation), (v) an agent promoting tissue regeneration *in vitro* and *in vivo*, and (vi) a biological defense weapon.

The MOL5 nucleic acids and proteins have applications in the diagnosis and/or treatment of various diseases and disorders. For example, the compositions of the present invention will have efficacy for the treatment of patients suffering from: Rheumatoid arthritis

(RA), CNS disorders, Alzheimer, Down syndrome, Schizophrenia, Parkinsons diseases as well as other diseases, disorders and conditions.

These materials are further useful in the generation of antibodies that bind immuno-specifically to the novel MOL5 substances for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-MOLX Antibodies" section below. The disclosed MOL5 protein has multiple hydrophilic regions, each of which can be used as an immunogen. In one embodiment, a contemplated MOL5 epitope is from about amino acids 30 to 70. In another embodiment, a MOL5 epitope is from about amino acids 100 to 150. In additional embodiments, MOL5 epitopes are from about amino acids 175 to 200, 220 to 450, 550 to 575, 590 to 610, and from about amino acids 675 to 850. These novel proteins can also be used to develop assay systems for functional analysis.

MOL6

The disclosed novel kappa casein precursor -like MOL6 nucleic acid of 603 nucleotides (also referred to as GMAC060288_A) is shown in Table 6A. An open reading begins with an ATG initiation codon at nucleotides 31-33 and ends with a TAA codon at nucleotides 574-576. A putative untranslated region upstream from the initiation codon and downstream from the termination codon are underlined in Table 6A, and the start and stop codons are in bold letters.

Table 6A. MOL6 Nucleotide Sequence (SEQ ID NO:19)

TTTTTTTTAAATTTATCTTTAGGTGCAATAATGAAGAGTTTTCTTCTAGTTGTCAATGCCCTGGCATTAACCCTG
CCTTTTTTGCTAGTGGAGGTTCAAAACCAGAAACAACCAGCATGCCATGAGAATGATGAAAGACCATTCTATCAG
AAAACGTTCACATATGTCCCAATGTATTATGTGCAAAATAGCTATCTTTATTATGGACCCAATTTGTACAAACGT
AGACCAGCTATAGCATTAAATAATCAATATGGGCTTCGCACATATTATGCAACCCAAGCTGTAGTTAGGGCACAT
GCCCAAATTCCTCAGCGGCAATACCTGCCAAATAGCCACCACACTGTGGTACGTCGCCCAAACCTGCATCCATCA
TTTATTGCAATCCCCCCAAAGAAAATTCAGGATAAAATAATCATCCCTACCATCAATACCATTGCTACTGTTGAA
CCTACACCAGCTCCTGCCACTGAACCAACGGTGGACAGTGTAATCACTCCAGAAGCTTTTTCAGAGTCCATCATC
ACGAGCACCCCTGAGACAACCACAGTTGCAGTTACTCCACCTACGGCATAAAAACACCAAGGAAATATCAAAGAA
CAC

The MOL6 protein encoded by SEQ ID NO:20 has 181 amino acid residues, and is presented using the one-letter code in Table 6B (SEQ ID NO:20). The Psort profile for MOL6 predicts that this sequence has a signal peptide and is likely to be localized outside the cell with a certainty of 0.8200. The most likely cleavage site for a peptide is between amino acids

24 and 25: VQN-QK based on the SignalP result. The molecular weight of the MOL6 protein is 20424.3 Daltons.

Table 6B. Encoded MOL6 protein sequence (SEQ ID NO:20).

```
MKSFLLVVNALALTLPFLLVEVQNQKQPACHENDERPFYQKTFTYVPMYYVQNSYLYYGPNLYKRRPAIALNNQYG
LRTYYATQAVVRAHAQIPQRQYLPNSHHTVVRRPNLHPSFIAIPPKKIQDKIIIPTINTIATVEPTPAPATEPTVD
SVITPEAFSESIITSTPETTTVAVTPPTA
```

The disclosed nucleic acid sequence has 566 of 586 bases (96 %) identical to a *Homo sapiens* kappa casein precursor mRNA (GENBANK-ID: ACC: I29004) (E value = $9.8e\text{-}^{116}$).

The full amino acid sequence of MOL6 was found to have 165 of 182 amino acid residues (90%) identical to, and 168 of 182 residues (92%) positive with, the 182 amino acid residue kappa casein precursor protein from *Homo sapiens* (ptnr: SWISSPROT-ACC:P07498) (E value = $3.0e\text{-}^{83}$), 165 of 182 amino acid residues (90%) identical to, and 168 of 182 residues (92%) positive with patp:AAR39351 Recombinant human kappa casein - *Homo sapiens* having 182 aa (E value = $3.0e\text{-}^{83}$), and 165 of 182 amino acid residues (90%) identical to, and 168 of 182 residues (92%) positive with patp:AAR92150 Human milk kappa-casein having 182 amino acids (E value = $3.0e\text{-}^{83}$).

The global sequence homology (as defined by FASTA alignment with the full length sequence of this protein) is 92.265 % amino acid homology and 91.160% amino acid identity. In addition, this protein contains the following protein domains (as defined by Interpro) at the indicated nucleotide positions: casein_kappa (IPR000117) at amino acid positions 1 to 181.

The full amino acid sequence of MOL6 was found to have homology with several proteins, including those disclosed in the BLASTP data in Table 6C.

Table 6C. BLAST results for MOL6

Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi|1705606|sp|P07498|CASK_HUMAN	KAPPA CASEIN PRECURSOR	182	154/171 (90%)	157/171 (91%)	3e-59
gi|4885161|ref|NP_005203.1|	casein, kappa [Homo sapiens]	182	153/171 (89%)	156/171 (90%)	9e-59
gi|186655|gb|AAA59456.1|	kappa-casein [Homo sapiens]	182	153/171 (89%)	156/171 (90%)	6e-55
gi|13633560|ref|XP_003538.3|	casein, kappa [Homo sapiens]	182	144/171 (84%)	147/171 (85%)	3e-54
gi|2493502|sp|P79139|CASK_CAMDR	KAPPA CASEIN PRECURSOR	182	102/178 (57%)	118/178 (65%)	2e-31

This information is presented graphically in the multiple sequence alignment given in Table 6D (with MOL6 being shown on line 1) as a ClustalW analysis comparing MOL6 with related protein sequences.

Table 6D Information for the ClustalW proteins:

1) MOL6 (SEQ ID NO:20)
2) gi|1705606|sp|P07498|CASK_HUMAN KAPPA CASEIN PRECURSOR (SEQ ID NO:50)
3) gi|4885161|ref|NP_005203.1| casein, kappa [Homo sapiens] (SEQ ID NO:51)
4) gi|186655|gb|AAA59456.1| kappa-casein [Homo sapiens] (SEQ ID NO:52)
5) gi|13633560|ref|XP_003538.3| casein, kappa [Homo sapiens] (SEQ ID NO:53)
6) gi|2493502|sp|P79139|CASK_CAMDR KAPPA CASEIN PRECURSOR (SEQ ID NO:54)

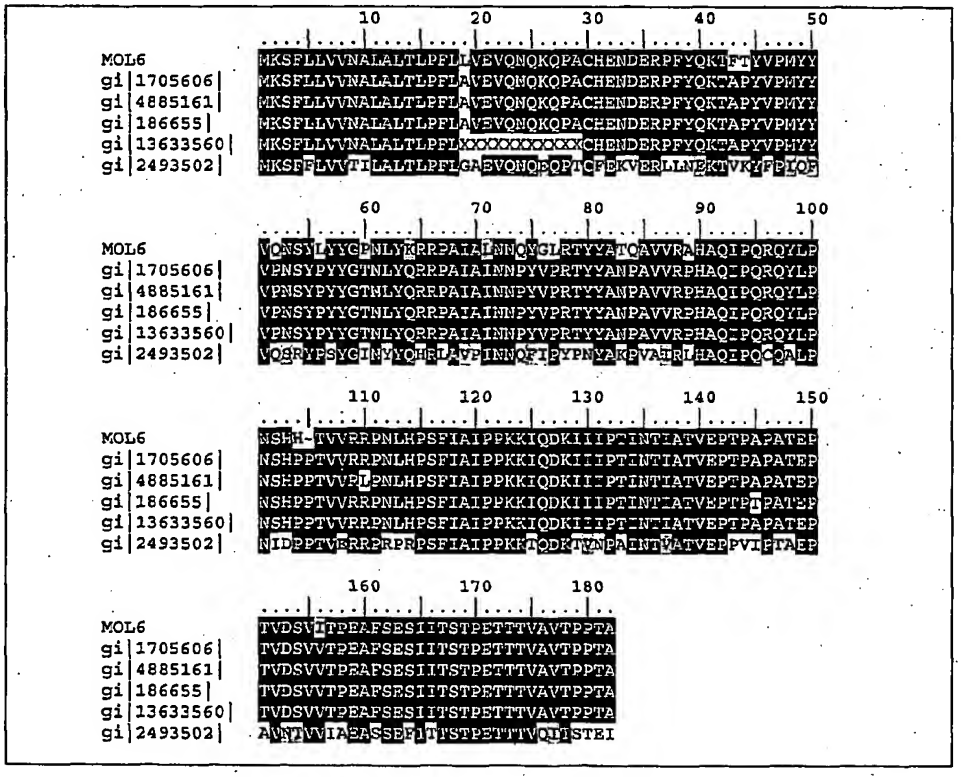
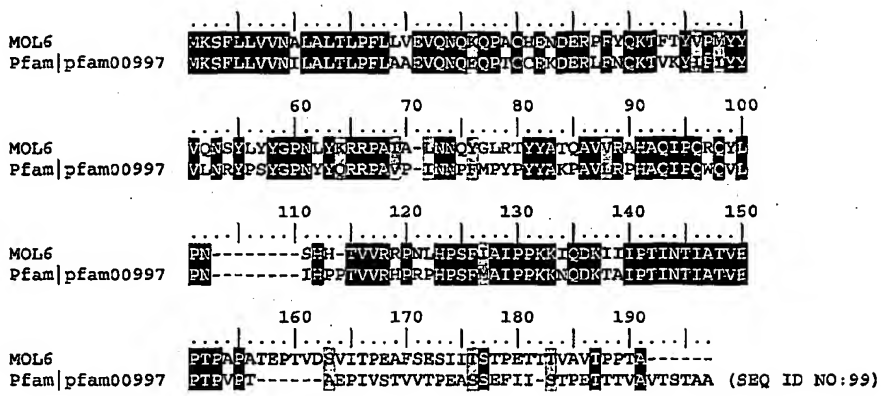


Table 6E lists the domain description from DOMAIN analysis results against MOL6. The region from amino acid residue 1 through 116 (SEQ ID NO:20) most probably ($E = 2e^{-36}$) contains a casein kappa domain found in Kappa casein, aligned here in Table 6E. This indicates that the MOL6 sequence has properties similar to those of other proteins known to contain this domain.

Table 6E. Domain Analysis of MOL6

gnl|Pfam|pfam00997, casein_kappa, Kappa casein
CD-Length = 181 residues, only 64.6% aligned
Score = 145 bits (366), Expect = 2e-36

10 20 30 40 50



(SEQ ID NO:99)

The above defined information for MOL6 suggests that this kappa casein precursor-like protein may function as a member of a "Kappa Casein Precursor family". Members of this family is found as a nutritional component of human milk. Therefore, the novel nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for MOL6 include, but are not limited to: protein therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, kappa casein precursor therapy (kappa casein precursor delivery/kappa casein precursor ablation), research tools, tissue reKappa Casein Precursor ration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

The MOL6 nucleic acids and proteins are useful in potential therapeutic applications implicated in nutritional deficiencies. It is used as a nutrient supplement in milk based products to provide a substantial improvement of the nutritional and biological value of the formulae, making it closer in similarity to human milk. Kappa casein can also be used as a pharmaceutical and/or other pathologies and disorders. For example, a cDNA encoding the kappa casein precursor-like protein may be useful in kappa casein precursor therapy, and the kappa casein precursor-like protein may be useful when administered to a subject in need thereof. By way of nonlimiting example, the compositions of the present invention will have efficacy for treatment of patients suffering from nutritional deficiencies. MOL6, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

These materials are further useful in the generation of antibodies that bind immunospecifically to the novel MOL6 substances for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from

hydrophobicity charts, as described in the "Anti-MOLX Antibodies" section below. The disclosed MOL6 protein has multiple hydrophilic regions, each of which can be used as an immunogen. In one embodiment, a contemplated MOL6 epitope is from about amino acids 30 to 125. In another embodiment, a MOL6 epitope is from about amino acids 140 to 160. These novel proteins can also be used to develop assay systems for functional analysis.

MOL7

A novel nucleic acid encoding a human Rh type B glycoprotein -like-protein MOL7 was identified by TblastN using CuraGen Corporation's sequence file for MOL7 probes or homologs, and run against the Genomic Daily Files made available by GenBank. The disclosed novel MOL7 nucleic acid of 1765 nucleotides (also referred to as AF193808A) is shown in Table 7A. An open reading frame begins with an ATG initiation codon at nucleotides 39-41 and ends with a TAA codon at nucleotides 1383-1385. A putative untranslated region upstream from the initiation codon and downstream from the termination codon are underlined in Table 7A, and the start and stop codons are in bold letters.

Table 7A. MOL7 Nucleotide Sequence (SEQ ID NO:21)

AAAGCCTGCGAGCGCCAGCCGAGATCGCATCCCAACCCATGGCCGGGTCTCCTAGCCGCGCCGCGGGCCGGCGACTGC
AGCTTCCCCTGCTGTGCCTCTTCCTCCAGGGCGCCACTGCCGTCCTCTTTGCTGTCTTTGTCCGCTACAACCACAAAA
CCGACGCTGCCCTCTGGCACCGGAGCAACCACAGTAACGCGGACAATGAATTTTACTTTCGCTACCCAAGTTTCCAGG
ACGTGCATGCCATGGTCTTCGTGGGCTTTGACTTCCTCATGGTCTTCCTGCAGCGTTACGGCTTCAGCAGCGTGGGCT
TCACCTTCCTCCTGGCCGCCTTTGCCCTGCAGTGGTCCACACTGGTCCAGGGCTTTCTCCACTCCTTCCACGGTGGCC
ACATCCATGTTGGCGTGGAGAGCATGATCAATGCTGACTTTTGTGCGGGGGCCGTGCTCATCTCCTTTGGTGCCGTCC
TGGGCAAGACCGGGCCTACCCAGCTGCTGCTCATGGCCCTGCTGGAGGTGGTGCTGTTTGGCATCAATGAGTTTGTGC
TCCTTCATCTCCTGGGGGTGAGAGTCTGGGGAGGGATTTCTAGGGTTATGTCTAGTACCATGCTGGAGAAGAGCAAGC
ACCGCCAGGGCTCCGTCTACCATTCAGACCTCTTCGCCATGATTGGTGGGACCATCTTCCTGTGGATCTTCTGGCCTA
GCTTCAATGCTGCACTCACAGCGCTGGGGGCTGGGCAGCATCGGACGGCCCTCAACACATACTACTCCCTGGCTGCCA
GCACCCTTGGCACCTTTGCCTTGTCAGCCCTTGTAGGGGAAGATGGGAGGCTTGACATGGTAGTCCACATCCAAAATG
CAGCGCTGGCTGGAGGGGTTGTGGTGGGGACCTCAAGTGAAATGATGCTGACACCCTTTGGGGCTCTGGCAGCTGGCT
TCTTGGCTGGGACTGTCTCCACGCTGGGGTACAAGTTCTTCACGCCCATCCTTGAATCAAAATTCAAAGTCCAAGACA
CATGTGGAGTCCACAACCTCCATGGGATGCCGGGGGTCCTGGGGGCCCTCCTGGGGGTCCTTGTGGCTGGACTTGCCA
CCCATGAAGCTTACGGAGATGGGCTGGAGAGTGTGTTTCCACTCATAGCCGAGGGCCAGCGCAGTGCCACGTCACAGG
CCATGCACCAGCTCTTCGGGCTGTTTGTCACACTGATGTTTGCCTCTGTGGGCGGGGGCCTTGGAGGTGGGCTCCTGC
TGAAGCTACCCTTTCTGGACTCCCCCCCCGACTCCCAGCACTACGAGGACCAAGTTCACTGGCAGGTGGTGCCTGGCG
AGCATGAGGATAAAGCCCAGAGACCTCTGAGGGTGGAGGAGGCAGACACTCAGGCCTAACCCACTGCCAGCCCCTGAG
AGGACACGCTCCTTTTCGAAGATGCTGACTGGCTGCTACTAGGAAGTTCTTTTTGAGCTCCCATTCCTCCAGCTGCAA
GAAGGGAGCCATGAGCCAGAAGGAGGCCCCTTTCCACAGGCAGCGTCTCCACAGGGAGAGGGGCAACAGGAGGCTGGG
AAATGGTGGGGAGTGGGGCCGTAACTGGGTACAATAGGGGGAACCTCACCAGATGCCCAACCCGACTGCCCTACCAGC
CTGCACATGGGTAGAAGAGGCCAAATTGAGGCACCCAAGTGATCCACTGGCCCCACGTCACACAGTTACAGTGAAGCC
CAAGCCAGGCCTGGTTGAGGGTGATAAACGCCACTGTCTTTAAGGAAAA

The MOL7 protein encoded by SEQ ID NO:21 has 448 amino acid residues, and is presented using the one-letter code in Table 7B (SEQ ID NO:22). The SignalP, Psort and/or Hydropathy profile for MOL7 predict that MOL7 has a signal peptide and is likely to be localized at the plasma membrane with a certainty of 0. 6400. The SignalP shows a signal sequence is coded for with the most likely cleavage site being between amino acids 27 and 28:

ATA-VL. This is typical of this type of membrane protein. The molecular weight of the MOL7 protein is 48304.3 Daltons.

Table 7B. Encoded MOL7 protein sequence (SEQ ID NO:22).

MAGSPSRAAGRRLQLPLLCLFLQGATAVLFAVFVRYNHKTDAALWHRSNHSNADNEFYFRYPSFQDV
HAMVFVGFDFLMVFLQRYGFSSVGFTFLLAAFALQWSTLVQGFLHSFHGGHIHVGVESMINADFCAG
AVLISFGAVLGKTGPTQLLLMALLEVVLFGINEFVLLHLLGVRVWGGISRVMSSTMLEKSKHRQGSV
YHSDLFAMIGGTIFLWIFWPSFNAALTALGAGQHRTALNTYYSLAASTLGTFALSALVGEDGRLDMV
VHIQNAALAGGVVVGTSSEMMLTPFGALAAGFLAGTVSTLGYKFFTPILESKFKVQDTCGVHNLHGM
PGVLGALLGVLVAGLATHEAYGDGLESVFPLIAEGQRSATSQAMHQLFGLFVTLMFASVGGGLGGGL
LLKLPFLDSPPDSQHYEDQVHWQVVPGEHEDKAQRPLRVEEADTQA

The nucleic acid sequence of MOL7 was found to have 680 of 815 bases (83) identical to a mouse Rh type b glycoprotein mRNA (GENBANK-ID:AF193808|acc:AF193808).

The full amino acid sequence of MOL7 was found to have 363 of 448 amino acid residues (81%) identical to, and 399 of 448 residues (89%) positive with, the 455 amino acid residue mouse RH TYPE B GLYCOPROTEIN (ptnr: SPTREMBL-ACC:Q9QXP1)

The full amino acid sequence of MOL7 was found to have homology with several proteins, including those disclosed in the BLASTP data in Table 7C.

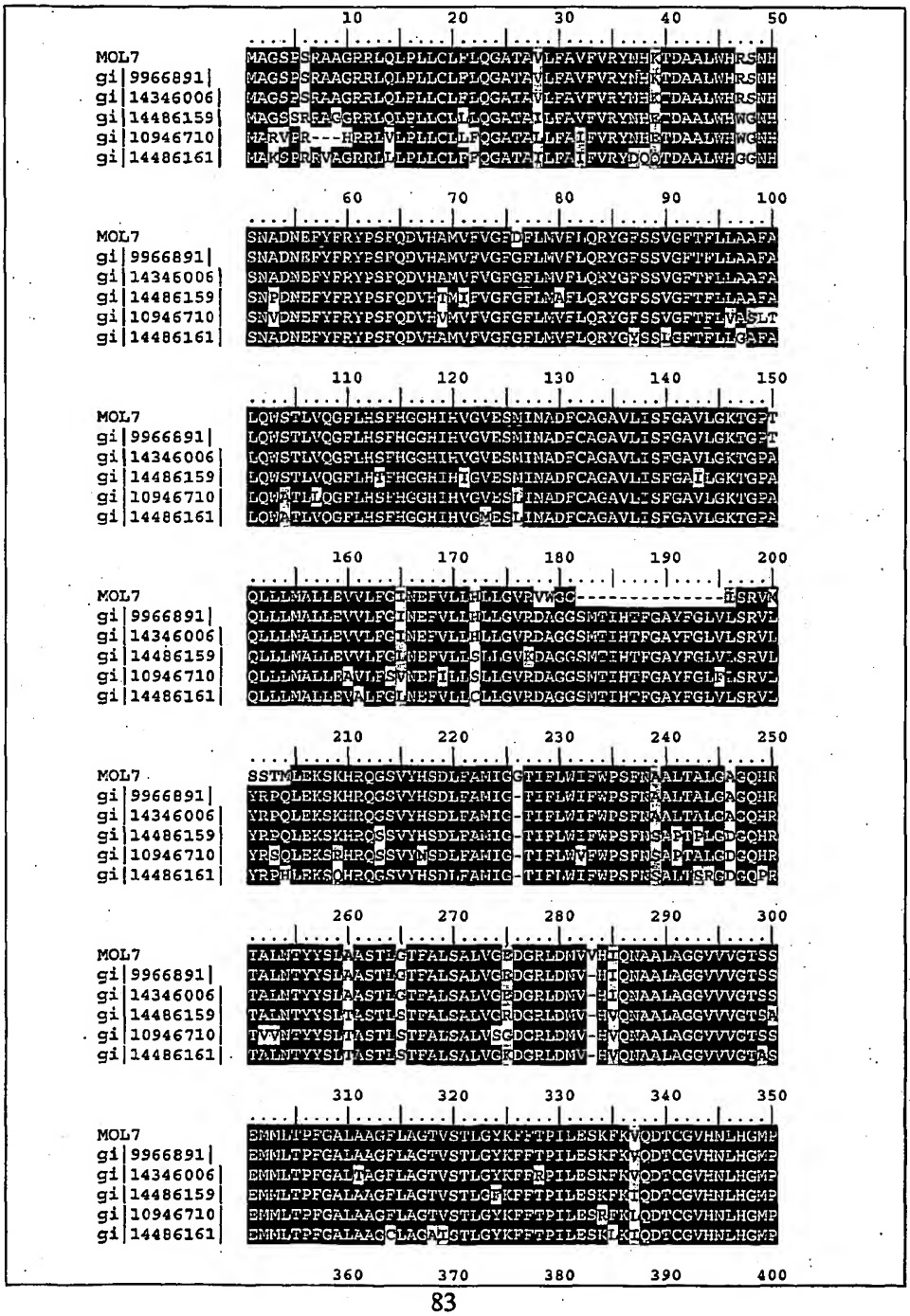
Table 7C. BLAST results for MOL7

Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi|9966891|ref|NP_065140.1|	Rh type B glycoprotein [Homo sapiens]	458	433/462 (93%)	435/462 (93%)	0.0
gi|14346006|gb|AAK15395.1| (AY013268)	Rh type B glycoprotein [Pan troglodytes]	458	429/462 (92%)	431/462 (92%)	0.0
gi|14486159|gb|AAK14651.1| (AY013261)	Rh type B glycoprotein [Sus scrofa]	458	384/462 (83%)	409/462 (88%)	0.0
gi|10946710|ref|NP_067350.1|	Rhesus blood group-associated B glycoprotein; Rh type B glycoprotein [Mus musculus]	455	362/452 (80%)	398/452 (87%)	e-177
gi|14486161|gb|AAK14652.1| (AY013262)	Rh type B glycoprotein [Oryctolagus cuniculus]	458	373/462 (80%)	404/462 (86%)	e-176

This information is presented graphically in the multiple sequence alignment given in Table 7D (with MOL7 being shown on line 1) as a ClustalW analysis comparing MOL7 with related protein sequences.

Table 7D Information for the ClustalW proteins:

1) MOL7 (SEQ ID NO:22)
2) gi|9966891|ref|NP_065140.1| h type B glycoprotein [Homo sapiens] (SEQ ID NO:55)
3) gi|14346006|gb|AAK15395.1| (AY013268) Rh type B glycoprotein [Pan troglodytes] (SEQ ID NO:56)
4) gi|14486159|gb|AAK14651.1| (AY013261) Rh type B glycoprotein [Sus scrofa] (SEQ ID NO:57)
5) gi|10946710|ref|NP_067350.1| Rhesus blood group-associated B glycoprotein; Rh type B glycoprotein [Mus musculus] (SEQ ID NO:58)
6) gi|14486161|gb|AAK14652.1| (AY013262) Rh type B glycoprotein [Oryctolagus cuniculus] (SEQ ID NO:59)



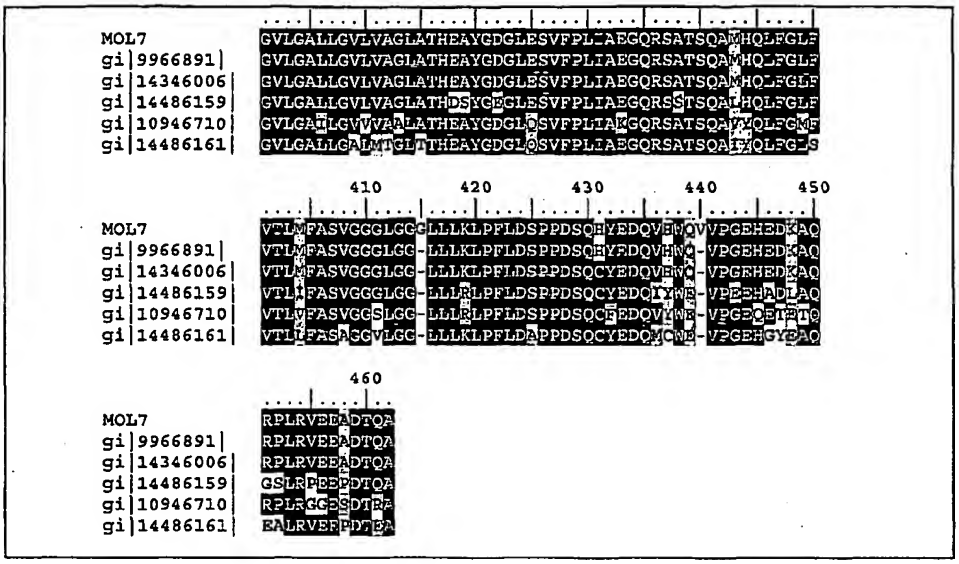
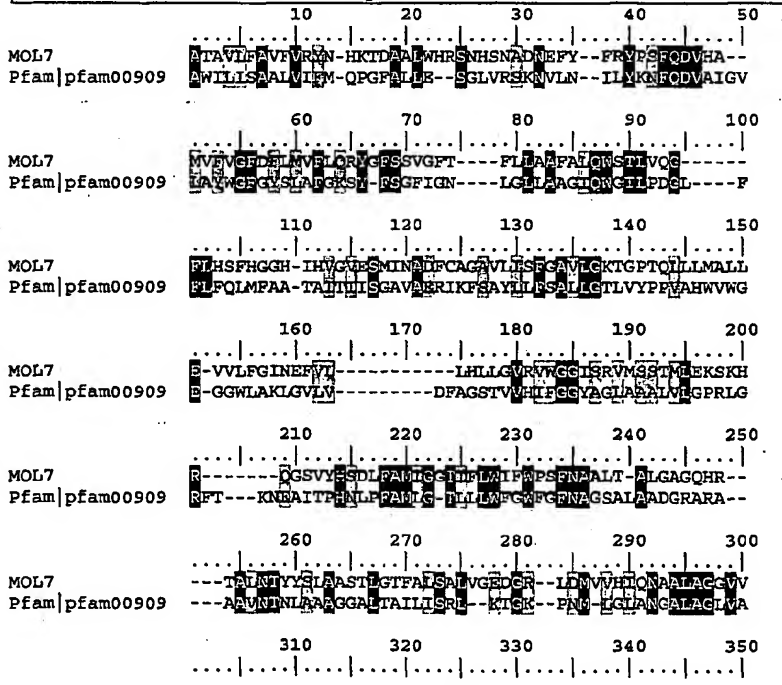
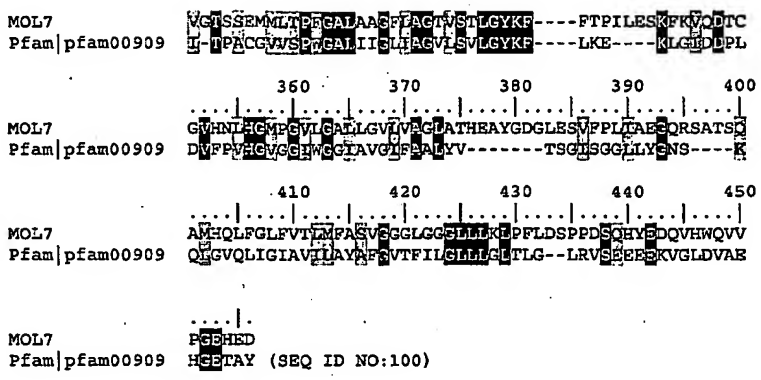


Table 7E lists the domain description from DOMAIN analysis results against MOL7. The region from amino acid residue 25 through 336 (SEQ ID NO:22) most probably ($E = 1e^{-33}$) contains an ammonium transporter domain found in Ammonium transporters, aligned here in Table 7E. This indicates that the MOL7 sequence has properties similar to those of other proteins known to contain this domain.

Table 7E. Domain Analysis of MOL7

gnl|Pfam|pfam00909, Ammonium_transp, Ammonium Transporter Family
CD-Length = 395 residues, only 78.7% aligned
Score = 137 bits (345), Expect = 1e-33





TaqMan Data

Example 2 shows a TaqMan expression profile in 41 normal human tissues and 55 human cancer cell lines. The MOL7 gene is expressed in normal tissues, specifically lung, colon, small intestine, and prostate, and is lost in cancer cell lines.

Example 2 also shows replicate TaqMan expression results in tumor tissues that are often matched with normal adjacent tissue (NAT), as defined by the operating surgeon. The results reveal that the MOL7 human Rh type B glycoprotein is overexpressed in kidney tumors compared with their NAT and normal tissues.

Chromosomal localization:

This gene belongs to genomic DNA GenBank AL139130 which maps to chromosome 1.

Tissue expression:

MOL7 has been found to be expressed in Renal clear cell carcinoma by EST analysis. Genbank EST AI310325 has 100% identity with novel Rh type B glycoprotein and was obtained from 2 pooled tumors (clear cell type). Kidney, AI925934 has 100% identity with novel Rh type B glycoprotein and was obtained from Kidney. Fetal spleen R83833 and AI022447 have 96% identity to novel Rh type B glycoprotein and were obtained from Fetal spleen. The tissue expression profile of was also determined by TaqMan.

Uses of the Compositions of the Invention

The expression pattern, map location and protein similarity information for the MOL7 suggest that this gene may function as "an Rh family" member. Therefore, the MOL7 nucleic acids and proteins are useful in potential therapeutic applications implicated in various pathologies /disorders described and/or other pathologies/disorders

Potential therapeutic uses for MOL7 include: Protein therapeutic, Small molecule drug target, Antibody target (Therapeutic, Diagnostic, Drug targeting/Cytotoxic antibody), Diagnostic and/or prognostic marker, Gene therapy (gene delivery/gene ablation), Research tools, Tissue regeneration *in vitro* and *in vivo* (regeneration for all these tissues and cell types composing these tissues and cell types derived from these tissues)

The MOL7 nucleic acids and proteins are useful in potential therapeutic applications implicated in various names of pathologies/disorders described below and/or other pathologies disorders. For example, a cDNA encoding the RH TYPE B GLYCOPROTEIN -like protein may be useful in gene therapy, and the RH TYPE B GLYCOPROTEIN-like protein may be useful when administered to a subject in need thereof. By way of nonlimiting example, the compositions of the present invention will have efficacy for treatment of patients suffering from the pathologies described above. The novel nucleic acid encoding MOL7, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

These materials are further useful in the generation of antibodies that bind immunospecifically to the novel MOL7 substances for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-MOLX Antibodies" section below. The disclosed MOL7 protein has multiple hydrophilic regions, each of which can be used as an immunogen. In one embodiment, a contemplated MOL7 epitope is from about amino acids 40 to 80. In another embodiment, a MOL7 epitope is from about amino acids 160 to 190. In additional embodiments, MOL7 epitopes are from about amino acids 175 to 225, 235 to 250, 325 to 345, 360 to 380, and from about amino acids 400 to 450. These novel proteins can also be used to develop assay systems for functional analysis.

MOL8

A novel human Noelin-2-like nucleic acid was identified by TblastN using CuraGen Corporation's sequence file. The disclosed novel MOL8 nucleic acid of 1399 nucleotides (also referred to as SC84366578_A) is shown in Table 8A. An open reading frame begins with an ATG initiation codon at nucleotides 14-16 and ends with a TAG codon at nucleotides 1391-1393. A putative untranslated region downstream from the termination codon are underlined in Table 8A, and the start and stop codons are in bold letters.

Table 8A. MOL8 Nucleotide Sequence (SEQ ID NO:23)

```
TGTTTTACTTGAAATGCTACAAACCAACACTCTTTTTATCCTAAAACAGGAGTCTGTGTTTTATGTTTCCCTTTGG
TTTCCTCAGACTCAGATTAGTCCTAAAGAAGGGTGGCAGGTGTACAGCTCAGCTCAGGATCCTGATGGGCGGTGCA
TTTGCACAGTTGTTGCTCCAGAACAAAACCTGTGTTCCCGGGATGCCAAAAGCAGGCAACTTCGCCAACTACTGGA
AAAGGTACAGAACATGTCCCAGTCTATTGAAGTCTTAAACTTGAGAACTCAGAGAGATTTCCAATATGTTTTAAAA
ATGGAAACCCAAATGAAAGGGCTGAAGGCAAAATTTCGGCAGATTGAAGATGATCGAAAGACACTTATGACCAAGC
ATTTTCAGCAGGAGTTGAAAGAGAAAATGGACGAGCTCCTGCCTTTGATCCCCGTGCTGGAACAGTGCAAAACAGA
TGCTAAGTTCATCACCCAGTTCAAGGAGGAAATAAGGAATCTGTCTGCTGTCCTCACTGGTATTCAGGAGGAAATT
GGTGCCTATGACTACGAGGAACTACACCAAAGAGTGCTGAGCTTGGAAACAAGACTTCGTGACTGCATGAAAAAGC
TATGTGGCAAACTGATGAAAATCACAGGCCCAGTTACAGTCAAGACATCTGGAACCCGATTTGGTGCTTGGATGAC
AGACCCTTTAGCATCTGAGAAAAACAACAGAGTATGGTACATGGACAGTTATACTAACAATAAAATTGTTCGTGAA
TACAAATCAATTGCAGACTTTGTCAGTGGGGCTGAATCAAGGACATACAACCTTCCTTTCAAGTGGGCAGGAACTA
ACCATGTTGTCTACAATGGCTCACTCTATTTTAACAAGTATCAGAGTAATATCATCATCAAATACAGCTTTGATAT
GGGGAGAGTGCTTGCCCAACGAAGCCTGGAGTATGCTGGTTTTCATAATGTTTACCCCTACACATGGGGTGGATTC
TCTGACATCGACCTAATGGCTGATGAAATCGGGCTGTGGGCTGTGTATGCAACTAACCAGAATGCAGGCAATATTG
TCATCAGCCAACTTAACCAAGATACCTTGGAGGTGATGAAGAGCTGGAGCACTGGCTACCCCAAGAGAAGTGCAGG
GGAATCTTTCATGATCTGTGGGACACTGTATGTCACCAACTCCCACTTAACTGGAGCCAAGGTGTATTATTCCTAT
TCCACCAAAACCTCCACATATGAGTACACAGACATTCCCTTCCATAACCAATACTTTCACATATCCATGCTTGACT
ACAATGCAAGAGATCGAGCTCTCTATGCCTGGAACAATGGCCACCAGGTGCTGTTCAATGTCACCCTTTTCCATAT
CATCAAGACAGAGGATGACACATAGGCAAAT
```

The MOL8 protein encoded by SEQ ID NO:23 has 459 amino acid residues, and is presented using the one-letter code in Table 8B (SEQ ID NO:24). The SignalP, Psort and/or Hydropathy profile for MOL8 predict that MOL8 has no signal peptide and is likely to be localized at the microbody (peroxisome) with a certainty of 0.5616. The molecular weight of the MOL8 protein is 53275.2 Daltons.

Table 8B. Encoded MOL8 protein sequence (SEQ ID NO:24).

```
MLQTNTLFILKQESVFYVSLWFPQTQISPKEGWQVYSSAQDPDGRCICTVVAPEQNLCSRDAKSRQLRQLLEKVQN
MSQSIEVLNLRTQRDFQYVLKMETQMKGLKAKFRQIEDDRKTLMTKHFQQELKEKMDELLPLIPVLEQCKTDAKFI
TQFKEEIRNLSAVLTGIQEEIGAYDYEELHQRVLSLETRLRDCMKKLCGKLMKITGPVTVKTSGTRFGAWMTDPLA
SEKNNRVWYMDSYTNNKIVREYKSIADFVSGAESRTYNLPFKWAGTNHVVYNGSLYFNKYQSNIIIKYSFDMGRVL
AQRSLEYAGFHNVYPYTWGGFSDIDLMADEIGLWAVYATNQNAGNIVISQLNQDTLEVMKSWSTGYPKRSAGESFM
ICGTLYVTNSHLTGAKVYYSYSTKTSTYEYTDIPFHNQYFHISMLDYNARDRALYAWNNGHQVLFNVTLFHIIKTE
DDT
```

The nucleotide sequence of MOL8 has 889 of 1286 bases (69%) identical to a *Gallus gallus* NOELIN-2 mRNA (GENBANK-ID: AF239804). The full amino acid sequence of the protein of the invention was found to have 288 of 448 amino acid residues (64%) identical to, and 367 of 448 residues (80%) positive with, the 457 amino acid residue NOELIN-2 protein from *Gallus gallus* (Chicken) (ptnr:SPTREMBL-ACC: AAF43715), and 439 of 459 amino acid residues (95%) identical to, and 442 of 459 residues (96%) positive with, the 458 amino acid residue patp:AAB74696 Human membrane associated protein MEMAP-2.

The global sequence homology (as defined by FASTA alignment with the full length sequence of this protein) is 74% amino acid homology and 65% amino acid identity. In addition, this protein contains the following protein domain (as defined by Pfam) at the

indicated nucleotide positions: Olfactomedin-like domain (PF02191) at amino acid positions 201 to 451.

The full amino acid sequence of MOL8 was found to have homology with several proteins including those disclosed in the BLASTP data in Table 8C.

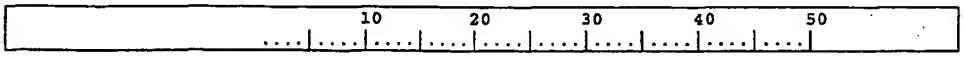
Table 8C. BLAST results for MOL8

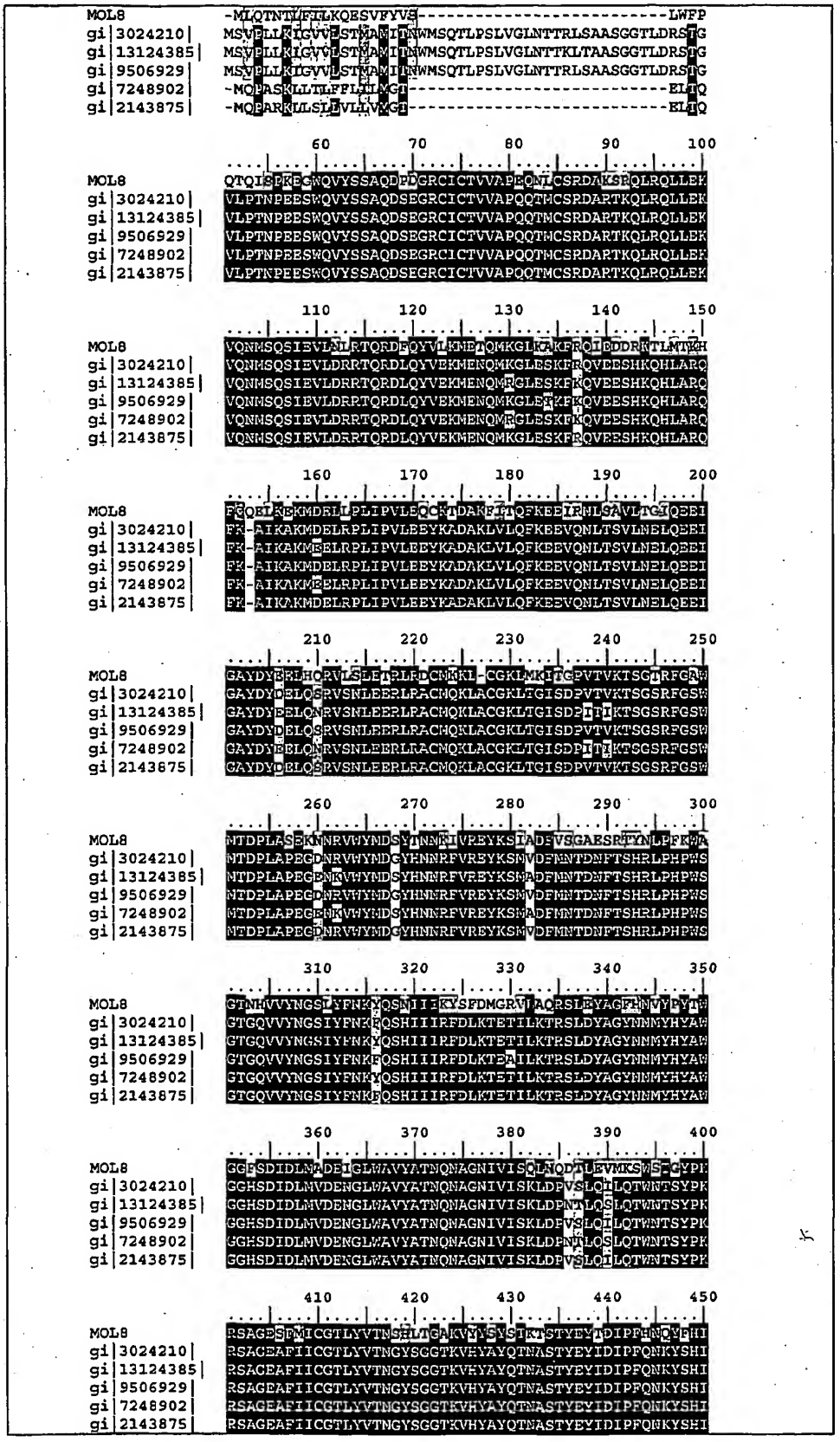
Gene Index/ Identifier	Protein/ Organism	Length (aa)	Identity (%)	Positives (%)	Expect
gi|3024210|sp|Q62609|NOE1_RAT	NOELIN PRECURSOR (NEURONAL OLFACTOMEDIN-RELATED ER LOCALIZED PROTEIN) (PANCORTIN) (1B426B)	485	285/431 (66%)	355/431 (82%)	e-172
gi|13124385|sp|Q9IAK4|NOE1_CHICK	NOELIN PRECURSOR (NEURONAL OLFACTOMEDIN-RELATED ER LOCALIZED PROTEIN) (PANCORTIN)	485	284/431 (65%)	357/431 (81%)	e-171
gi|9506929|ref|NP_062371.1|	olfactomedin related ER localized protein [Mus musculus]	485	284/431 (65%)	354/431 (81%)	e-171
gi|7248902|gb|AAF43715.1|AF239804_1 (AF239804)	NOELIN-2 [Gallus gallus]	457	284/431 (65%),	357/431 (81%)	e-171
gi|2143875|pir||I73636	neuronal olfactomedin-related ER localized protein - rat	457	285/431 (66%)	355/431 (82%)	e-171

Homology between MOL8 and other proteins are presented graphically in the multiple sequence alignment given in Table 8D (with MOL8 being shown on line 1) as a ClustalW analysis comparing MOL8 with related protein sequences.

Table 8D. Information for the ClustalW proteins:

1) MOL8 (SEQ ID NO:24)
2) gi|3024210|sp|Q62609|NOE1_RAT NOELIN PRECURSOR (NEURONAL OLFACTOMEDIN-RELATED ER LOCALIZED PROTEIN) (PANCORTIN) (1B426B) (SEQ ID NO:60)
3) gi|13124385|sp|Q9IAK4|NOE1_CHICK NOELIN PRECURSOR (NEURONAL OLFACTOMEDIN-RELATED ER LOCALIZED PROTEIN) (PANCORTIN) (SEQ ID NO:61)
4) gi|9506929|ref|NP_062371.1| olfactomedin related ER localized protein [Mus musculus] (SEQ ID NO:62)
5) gi|7248902|gb|AAF43715.1|AF239804_1 (AF239804) NOELIN-2 [Gallus gallus] (SEQ ID NO:63)
6) gi|2143875|pir||I73636 neuronal olfactomedin-related ER localized protein - rat (SEQ ID NO:64)




MOL8
gi|3024210|
gi|13124385|
gi|9506929|
gi|7248902|
gi|2143875|
60 70 80 90 100
110 120 130 140 150
160 170 180 190 200
210 220 230 240 250
260 270 280 290 300
310 320 330 340 350
360 370 380 390 400
410 420 430 440 450

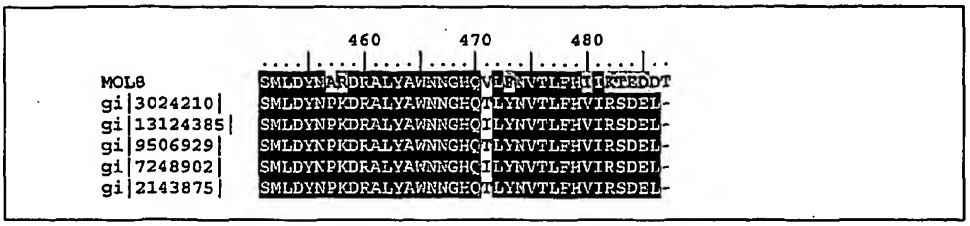
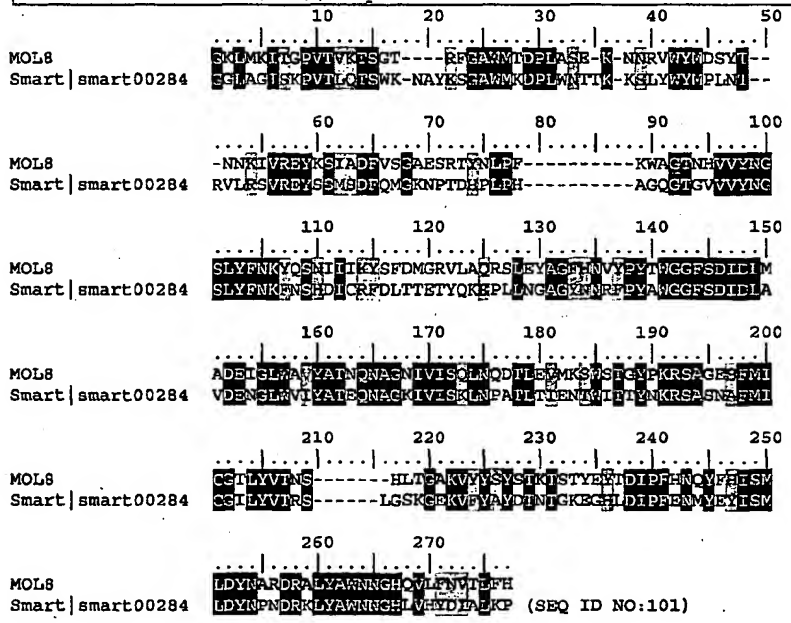


Table 8E lists the domain description from DOMAIN analysis results against MOL8. The region from amino acid residue 201 through 457 (SEQ ID NO:24) most probably ($E = 4e^{-85}$) contains a Olfactomedin-like domain, aligned in Table 8E. This indicates that the MOL8 sequence has properties similar to those of other proteins known to contain this domain.

Table 8E. Domain Analysis of MOL8

gnl|Smart|smart00284, OLF, Olfactomedin-like domains
CD-Length = 257 residues, 100.0% aligned
Score = 308 bits (789), Expect = 4e-85



Uses of the Compositions of the Invention

The above defined information for MOL8 suggests that this Noelin-2-like protein may function as a member of a "Noelin-2 family". This family is involved in neural crest development, and other developmental processes. Therefore, the novel nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for MOL8 include, but are not limited to: protein therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic

and/or prognostic marker, gene therapy (gene delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

The MOL8 nucleic acids and proteins are useful in potential therapeutic applications implicated in neural crest development in early embryonic stage. For example, a cDNA encoding the Noelin-2-like protein may be useful in gene therapy, and the Noelin-2-like protein may be useful when administered to a subject in need thereof. By way of nonlimiting example, the compositions of the present invention will have efficacy for treatment of patients suffering from primary open-angle glaucoma (POAG), and bone disorders, hematopoietic disorders, neuro-developmental disorders, cancer, autoimmune disorders, psychiatric disorders. The novel nucleic acid encoding MOL8, or fragments thereof, may further be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed.

These materials are further useful in the generation of antibodies that bind immuno-specifically to the novel MOL8 substances for use in therapeutic or diagnostic methods. These antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the "Anti-MOLX Antibodies" section below. The disclosed MOL8 protein has multiple hydrophilic regions, each of which can be used as an immunogen. In one embodiment, a contemplated MOL8 epitope is from about amino acids 20 to 50. In another embodiment, a MOL8 epitope is from about amino acids 50 to 125. In additional embodiments, MOL8 epitopes are from about amino acids 140 to 210, 225 to 320, 350 to 375, and from about amino acids 380 to 440. These novel proteins can also be used to develop assay systems for functional analysis.

TABLE 9: Summary Of Nucleic Acids And Proteins Of The Invention

Name	Tables	Clone; Description of Homolog	Nucleic Acid SEQ ID NO	Amino Acid SEQ ID NO
MOL1	1A, 1B,	MOL1a: SC29674552_EXT	1	2
	1D, 1E	MOL1b: CG56250-02	3	4
MOL2	2A, 2B	MOL2: SC98428706_EXT	5	6
MOL3	3A, 3B	MOL3a: SC85516573_EXT	7	8
	3D, 3E	MOL3b: CG53027-02	9	10
MOL4	4A, 4B,	MOL4: SC_111750277_A	11	12
MOL5	5A, 5B,	MOL5a: SC20422974_A	13	14
	5C, 5D,	MOL5b: SC14998905_EXT	15	16
	5E, 5F	MOLc: CG50907-02	17	18
MOL6	6A, 6B	MOL6: GMAC060288_A	19	20
MOL7	7A, 7B	MOL7: AF193808A	21	22
MOL8	8A, 8B	MOL8: SC84366578_A	23	24

MOLX Nucleic Acids and Polypeptides

One aspect of the invention pertains to isolated nucleic acid molecules that encode MOLX polypeptides or biologically active portions thereof. Also included in the invention are nucleic acid fragments sufficient for use as hybridization probes to identify MOLX-encoding nucleic acids (*e.g.*, MOLX mRNAs) and fragments for use as PCR primers for the amplification and/or mutation of MOLX nucleic acid molecules. As used herein, the term "nucleic acid molecule" is intended to include DNA molecules (*e.g.*, cDNA or genomic DNA), RNA molecules (*e.g.*, mRNA), analogs of the DNA or RNA generated using nucleotide analogs, and derivatives, fragments and homologs thereof. The nucleic acid molecule may be single-stranded or double-stranded, but preferably is comprised double-stranded DNA.

An MOLX nucleic acid can encode a mature MOLX polypeptide. As used herein, a "mature" form of a polypeptide or protein disclosed in the present invention is the product of a naturally occurring polypeptide or precursor form or proprotein. The naturally occurring polypeptide, precursor or proprotein includes, by way of nonlimiting example, the full-length gene product, encoded by the corresponding gene. Alternatively, it may be defined as the polypeptide, precursor or proprotein encoded by an ORF described herein. The product "mature" form arises, again by way of nonlimiting example, as a result of one or more naturally occurring processing steps as they may take place within the cell, or host cell, in which the gene product arises. Examples of such processing steps leading to a "mature" form of a polypeptide or protein include the cleavage of the N-terminal methionine residue encoded

by the initiation codon of an ORF, or the proteolytic cleavage of a signal peptide or leader sequence. Thus a mature form arising from a precursor polypeptide or protein that has residues 1 to N, where residue 1 is the N-terminal methionine, would have residues 2 through N remaining after removal of the N-terminal methionine. Alternatively, a mature form arising from a precursor polypeptide or protein having residues 1 to N, in which an N-terminal signal sequence from residue 1 to residue M is cleaved, would have the residues from residue M+1 to residue N remaining. Further as used herein, a "mature" form of a polypeptide or protein may arise from a step of post-translational modification other than a proteolytic cleavage event. Such additional processes include, by way of non-limiting example, glycosylation, myristoylation or phosphorylation. In general, a mature polypeptide or protein may result from the operation of only one of these processes, or a combination of any of them.

The term "probes", as utilized herein, refers to nucleic acid sequences of variable length, preferably between at least about 10 nucleotides (nt), 100 nt, or as many as approximately, *e.g.*, 6,000 nt, depending upon the specific use. Probes are used in the detection of identical, similar, or complementary nucleic acid sequences. Longer length probes are generally obtained from a natural or recombinant source, are highly specific, and much slower to hybridize than shorter-length oligomer probes. Probes may be single- or double-stranded and designed to have specificity in PCR, membrane-based hybridization technologies, or ELISA-like technologies.

The term "isolated" nucleic acid molecule, as utilized herein, is one, which is separated from other nucleic acid molecules which are present in the natural source of the nucleic acid. Preferably, an "isolated" nucleic acid is free of sequences which naturally flank the nucleic acid (*i.e.*, sequences located at the 5'- and 3'-termini of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. For example, in various embodiments, the isolated MOLX nucleic acid molecules can contain less than about 5 kb, 4 kb, 3 kb, 2 kb, 1 kb, 0.5 kb or 0.1 kb of nucleotide sequences which naturally flank the nucleic acid molecule in genomic DNA of the cell/tissue from which the nucleic acid is derived (*e.g.*, brain, heart, liver, spleen, etc.). Moreover, an "isolated" nucleic acid molecule, such as a cDNA molecule, can be substantially free of other cellular material or culture medium when produced by recombinant techniques, or of chemical precursors or other chemicals when chemically synthesized.

A nucleic acid molecule of the invention, *e.g.*, a nucleic acid molecule having the nucleotide sequence of SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, and 23, or a complement of this aforementioned nucleotide sequence, can be isolated using standard

molecular biology techniques and the sequence information provided herein. Using all or a portion of the nucleic acid sequence of SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, and 23 as a hybridization probe, MOLX molecules can be isolated using standard hybridization and cloning techniques (*e.g.*, as described in Sambrook, *et al.*, (eds.), MOLECULAR CLONING: A LABORATORY MANUAL 2nd Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989; and Ausubel, *et al.*, (eds.), CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, New York, NY, 1993.)

A nucleic acid of the invention can be amplified using cDNA, mRNA or alternatively, genomic DNA, as a template and appropriate oligonucleotide primers according to standard PCR amplification techniques. The nucleic acid so amplified can be cloned into an appropriate vector and characterized by DNA sequence analysis. Furthermore, oligonucleotides corresponding to MOLX nucleotide sequences can be prepared by standard synthetic techniques, *e.g.*, using an automated DNA synthesizer.

As used herein, the term "oligonucleotide" refers to a series of linked nucleotide residues, which oligonucleotide has a sufficient number of nucleotide bases to be used in a PCR reaction. A short oligonucleotide sequence may be based on, or designed from, a genomic or cDNA sequence and is used to amplify, confirm, or reveal the presence of an identical, similar or complementary DNA or RNA in a particular cell or tissue. Oligonucleotides comprise portions of a nucleic acid sequence having about 10 nt, 50 nt, or 100 nt in length, preferably about 15 nt to 30 nt in length. In one embodiment of the invention, an oligonucleotide comprising a nucleic acid molecule less than 100 nt in length would further comprise at least 6 contiguous nucleotides of SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, and 23, or a complement thereof. Oligonucleotides may be chemically synthesized and may also be used as probes.

In another embodiment, an isolated nucleic acid molecule of the invention comprises a nucleic acid molecule that is a complement of the nucleotide sequence shown in SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, and 23, or a portion of this nucleotide sequence (*e.g.*, a fragment that can be used as a probe or primer or a fragment encoding a biologically-active portion of an MOLX polypeptide). A nucleic acid molecule that is complementary to the nucleotide sequence shown in SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, and 23 is one that is sufficiently complementary to the nucleotide sequence shown in SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, and 23 that it can hydrogen bond with little or no mismatches to the nucleotide sequence shown SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, and 23, thereby forming a stable duplex.

As used herein, the term "complementary" refers to Watson-Crick or Hoogsteen base pairing between nucleotides units of a nucleic acid molecule, and the term "binding" means the physical or chemical interaction between two polypeptides or compounds or associated polypeptides or compounds or combinations thereof. Binding includes ionic, non-ionic, van der Waals, hydrophobic interactions, and the like. A physical interaction can be either direct or indirect. Indirect interactions may be through or due to the effects of another polypeptide or compound. Direct binding refers to interactions that do not take place through, or due to, the effect of another polypeptide or compound, but instead are without other substantial chemical intermediates.

Fragments provided herein are defined as sequences of at least 6 (contiguous) nucleic acids or at least 4 (contiguous) amino acids, a length sufficient to allow for specific hybridization in the case of nucleic acids or for specific recognition of an epitope in the case of amino acids, respectively, and are at most some portion less than a full length sequence. Fragments may be derived from any contiguous portion of a nucleic acid or amino acid sequence of choice. Derivatives are nucleic acid sequences or amino acid sequences formed from the native compounds either directly or by modification or partial substitution. Analogs are nucleic acid sequences or amino acid sequences that have a structure similar to, but not identical to, the native compound but differs from it in respect to certain components or side chains. Analogs may be synthetic or from a different evolutionary origin and may have a similar or opposite metabolic activity compared to wild type. Homologs are nucleic acid sequences or amino acid sequences of a particular gene that are derived from different species.

Derivatives and analogs may be full length or other than full length, if the derivative or analog contains a modified nucleic acid or amino acid, as described below. Derivatives or analogs of the nucleic acids or proteins of the invention include, but are not limited to, molecules comprising regions that are substantially homologous to the nucleic acids or proteins of the invention, in various embodiments, by at least about 70%, 80%, or 95% identity (with a preferred identity of 80-95%) over a nucleic acid or amino acid sequence of identical size or when compared to an aligned sequence in which the alignment is done by a computer homology program known in the art, or whose encoding nucleic acid is capable of hybridizing to the complement of a sequence encoding the aforementioned proteins under stringent, moderately stringent, or low stringent conditions. See *e.g.* Ausubel, *et al.*, CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, New York, NY, 1993, and below.

A "homologous nucleic acid sequence" or "homologous amino acid sequence," or variations thereof, refer to sequences characterized by a homology at the nucleotide level or

amino acid level as discussed above. Homologous nucleotide sequences encode those sequences coding for isoforms of MOLX polypeptides. Isoforms can be expressed in different tissues of the same organism as a result of, for example, alternative splicing of RNA. Alternatively, isoforms can be encoded by different genes. In the invention, homologous nucleotide sequences include nucleotide sequences encoding for an MOLX polypeptide of species other than humans, including, but not limited to: vertebrates, and thus can include, *e.g.*, frog, mouse, rat, rabbit, dog, cat cow, horse, and other organisms. Homologous nucleotide sequences also include, but are not limited to, naturally occurring allelic variations and mutations of the nucleotide sequences set forth herein. A homologous nucleotide sequence does not, however, include the exact nucleotide sequence encoding human MOLX protein. Homologous nucleic acid sequences include those nucleic acid sequences that encode conservative amino acid substitutions (see below) in SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, and 23, as well as a polypeptide possessing MOLX biological activity. Various biological activities of the MOLX proteins are described below.

An MOLX polypeptide is encoded by the open reading frame ("ORF") of an MOLX nucleic acid. An ORF corresponds to a nucleotide sequence that could potentially be translated into a polypeptide. A stretch of nucleic acids comprising an ORF is uninterrupted by a stop codon. An ORF that represents the coding sequence for a full protein begins with an ATG "start" codon and terminates with one of the three "stop" codons, namely, TAA, TAG, or TGA. For the purposes of this invention, an ORF may be any part of a coding sequence, with or without a start codon, a stop codon, or both. For an ORF to be considered as a good candidate for coding for a *bona fide* cellular protein, a minimum size requirement is often set, *e.g.*, a stretch of DNA that would encode a protein of 50 amino acids or more.

The nucleotide sequences determined from the cloning of the human MOLX genes allows for the generation of probes and primers designed for use in identifying and/or cloning MOLX homologues in other cell types, *e.g.* from other tissues, as well as MOLX homologues from other vertebrates. The probe/primer typically comprises substantially purified oligonucleotide. The oligonucleotide typically comprises a region of nucleotide sequence that hybridizes under stringent conditions to at least about 12, 25, 50, 100, 150, 200, 250, 300, 350 or 400 consecutive sense strand nucleotide sequence of SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, and 23; or an anti-sense strand nucleotide sequence of SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, and 23; or of a naturally occurring mutant of SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, and 23.

Probes based on the human MOLX nucleotide sequences can be used to detect transcripts or genomic sequences encoding the same or homologous proteins. In various embodiments, the probe further comprises a label group attached thereto, *e.g.* the label group can be a radioisotope, a fluorescent compound, an enzyme, or an enzyme co-factor. Such probes can be used as a part of a diagnostic test kit for identifying cells or tissues which mis-express an MOLX protein, such as by measuring a level of an MOLX-encoding nucleic acid in a sample of cells from a subject *e.g.*, detecting MOLX mRNA levels or determining whether a genomic MOLX gene has been mutated or deleted.

"A polypeptide having a biologically-active portion of an MOLX polypeptide" refers to polypeptides exhibiting activity similar, but not necessarily identical to, an activity of a polypeptide of the invention, including mature forms, as measured in a particular biological assay, with or without dose dependency. A nucleic acid fragment encoding a "biologically-active portion of MOLX" can be prepared by isolating a portion SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, and 23 that encodes a polypeptide having an MOLX biological activity (the biological activities of the MOLX proteins are described below), expressing the encoded portion of MOLX protein (*e.g.*, by recombinant expression *in vitro*) and assessing the activity of the encoded portion of MOLX.

MOLX Nucleic Acid and Polypeptide Variants

The invention further encompasses nucleic acid molecules that differ from the nucleotide sequences shown SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, and 23 due to degeneracy of the genetic code and thus encode the same MOLX proteins as that encoded by the nucleotide sequences shown in SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, and 23. In another embodiment, an isolated nucleic acid molecule of the invention has a nucleotide sequence encoding a protein having an amino acid sequence shown in SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, and 24.

In addition to the human MOLX nucleotide sequences shown in SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, and 23 it will be appreciated by those skilled in the art that DNA sequence polymorphisms that lead to changes in the amino acid sequences of the MOLX polypeptides may exist within a population (*e.g.*, the human population). Such genetic polymorphism in the MOLX genes may exist among individuals within a population due to natural allelic variation. As used herein, the terms "gene" and "recombinant gene" refer to nucleic acid molecules comprising an open reading frame (ORF) encoding an MOLX protein, preferably a vertebrate MOLX protein. Such natural allelic variations can typically result in

1-5% variance in the nucleotide sequence of the MOLX genes. Any and all such nucleotide variations and resulting amino acid polymorphisms in the MOLX polypeptides, which are the result of natural allelic variation and that do not alter the functional activity of the MOLX polypeptides, are intended to be within the scope of the invention.

Moreover, nucleic acid molecules encoding MOLX proteins from other species, and thus that have a nucleotide sequence that differs from the human sequence SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, and 23 are intended to be within the scope of the invention. Nucleic acid molecules corresponding to natural allelic variants and homologues of the MOLX cDNAs of the invention can be isolated based on their homology to the human MOLX nucleic acids disclosed herein using the human cDNAs, or a portion thereof, as a hybridization probe according to standard hybridization techniques under stringent hybridization conditions.

Accordingly, in another embodiment, an isolated nucleic acid molecule of the invention is at least 6 nucleotides in length and hybridizes under stringent conditions to the nucleic acid molecule comprising the nucleotide sequence of SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, and 23. In another embodiment, the nucleic acid is at least 10, 25, 50, 100, 250, 500, 750, 1000, 1500, or 2000 or more nucleotides in length. In yet another embodiment, an isolated nucleic acid molecule of the invention hybridizes to the coding region. As used herein, the term "hybridizes under stringent conditions" is intended to describe conditions for hybridization and washing under which nucleotide sequences at least 60% homologous to each other typically remain hybridized to each other.

Homologs (*i.e.*, nucleic acids encoding MOLX proteins derived from species other than human) or other related sequences (*e.g.*, paralogs) can be obtained by low, moderate or high stringency hybridization with all or a portion of the particular human sequence as a probe using methods well known in the art for nucleic acid hybridization and cloning.

As used herein, the phrase "stringent hybridization conditions" refers to conditions under which a probe, primer or oligonucleotide will hybridize to its target sequence, but to no other sequences. Stringent conditions are sequence-dependent and will be different in different circumstances. Longer sequences hybridize specifically at higher temperatures than shorter sequences. Generally, stringent conditions are selected to be about 5 °C lower than the thermal melting point (Tm) for the specific sequence at a defined ionic strength and pH. The Tm is the temperature (under defined ionic strength, pH and nucleic acid concentration) at which 50% of the probes complementary to the target sequence hybridize to the target sequence at equilibrium. Since the target sequences are generally present at excess, at Tm, 50% of the probes are occupied at equilibrium. Typically, stringent conditions will be those in

which the salt concentration is less than about 1.0 M sodium ion, typically about 0.01 to 1.0 M sodium ion (or other salts) at

pH 7.0 to 8.3 and the temperature is at least about 30°C for short probes, primers or oligonucleotides (*e.g.*, 10 nt to 50 nt) and at least about 60°C for longer probes, primers and oligonucleotides. Stringent conditions may also be achieved with the addition of destabilizing agents, such as formamide.

Stringent conditions are known to those skilled in the art and can be found in Ausubel, *et al.*, (eds.), CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6. Preferably, the conditions are such that sequences at least about 65%, 70%, 75%, 85%, 90%, 95%, 98%, or 99% homologous to each other typically remain hybridized to each other. A non-limiting example of stringent hybridization conditions are hybridization in a high salt buffer comprising 6X SSC, 50 mM Tris-HCl (pH 7.5), 1 mM EDTA, 0.02% PVP, 0.02% Ficoll, 0.02% BSA, and 500 mg/ml denatured salmon sperm DNA at 65°C, followed by one or more washes in 0.2X SSC, 0.01% BSA at 50°C. An isolated nucleic acid molecule of the invention that hybridizes under stringent conditions to the sequences of SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, and 23 corresponds to a naturally-occurring nucleic acid molecule. As used herein, a "naturally-occurring" nucleic acid molecule refers to an RNA or DNA molecule having a nucleotide sequence that occurs in nature (*e.g.*, encodes a natural protein).

In a second embodiment, a nucleic acid sequence that is hybridizable to the nucleic acid molecule comprising the nucleotide sequence of SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, and 23 or fragments, analogs or derivatives thereof, under conditions of moderate stringency is provided. A non-limiting example of moderate stringency hybridization conditions are hybridization in 6X SSC, 5X Denhardt's solution, 0.5% SDS and 100 mg/ml denatured salmon sperm DNA at 55°C, followed by one or more washes in 1X SSC, 0.1% SDS at 37°C. Other conditions of moderate stringency that may be used are well-known within the art. See, *e.g.*, Ausubel, et *al.* (eds.), 1993, CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, NY, and Kriegler, 1990; GENE TRANSFER AND EXPRESSION, A LABORATORY MANUAL, Stockton Press, NY.

In a third embodiment, a nucleic acid that is hybridizable to the nucleic acid molecule comprising the nucleotide sequences of SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, and 23 or fragments, analogs or derivatives thereof, under conditions of low stringency, is provided. A non-limiting example of low stringency hybridization conditions are hybridization in 35% formamide, 5X SSC, 50 mM Tris-HCl (pH 7.5), 5 mM EDTA, 0.02%

PVP, 0.02% Ficoll, 0.2% BSA, 100 mg/ml denatured salmon sperm DNA, 10% (wt/vol) dextran sulfate at 40°C, followed by one or more washes in 2X SSC, 25 mM Tris-HCl (pH 7.4), 5 mM EDTA, and 0.1% SDS at 50°C. Other conditions of low stringency that may be used are well known in the art (*e.g.*, as employed for cross-species hybridizations). See, *e.g.*, Ausubel, *et al.* (eds.), 1993, CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, NY, and Kriegler, 1990, GENE TRANSFER AND EXPRESSION, A LABORATORY MANUAL, Stockton Press, NY; Shilo and Weinberg, 1981. *Proc Natl Acad Sci USA* **78**: 6789-6792.

Conservative Mutations

In addition to naturally-occurring allelic variants of MOLX sequences that may exist in the population, the skilled artisan will further appreciate that changes can be introduced by mutation into the nucleotide sequences of SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, and 23 thereby leading to changes in the amino acid sequences of the encoded MOLX proteins, without altering the functional ability of said MOLX proteins. For example, nucleotide substitutions leading to amino acid substitutions at "non-essential" amino acid residues can be made in the sequence of SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, and 24. A "non-essential" amino acid residue is a residue that can be altered from the wild-type sequences of the MOLX proteins without altering their biological activity, whereas an "essential" amino acid residue is required for such biological activity. For example, amino acid residues that are conserved among the MOLX proteins of the invention are predicted to be particularly non-amenable to alteration. Amino acids for which conservative substitutions can be made are well-known within the art.

Another aspect of the invention pertains to nucleic acid molecules encoding MOLX proteins that contain changes in amino acid residues that are not essential for activity. Such MOLX proteins differ in amino acid sequence from SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, and 24 yet retain biological activity. In one embodiment, the isolated nucleic acid molecule comprises a nucleotide sequence encoding a protein, wherein the protein comprises an amino acid sequence at least about 45% homologous to the amino acid sequences of SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, and 24. Preferably, the protein encoded by the nucleic acid molecule is at least about 60% homologous to SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, and 24; more preferably at least about 70% homologous to SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, and 24; still more preferably at least about 80% homologous to SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, and 24; even more preferably at least about 90% homologous to SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, and 24; and most

preferably at least about 95% homologous to SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, and 24.

An isolated nucleic acid molecule encoding an MOLX protein homologous to the protein of SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, and 24 can be created by introducing one or more nucleotide substitutions, additions or deletions into the nucleotide sequence of SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, and 23 such that one or more amino acid substitutions, additions or deletions are introduced into the encoded protein.

Mutations can be introduced into SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, and 24 by standard techniques, such as site-directed mutagenesis and PCR-mediated mutagenesis. Preferably, conservative amino acid substitutions are made at one or more predicted, non-essential amino acid residues. A "conservative amino acid substitution" is one in which the amino acid residue is replaced with an amino acid residue having a similar side chain. Families of amino acid residues having similar side chains have been defined within the art. These families include amino acids with basic side chains (*e.g.*, lysine, arginine, histidine), acidic side chains (*e.g.*, aspartic acid, glutamic acid), uncharged polar side chains (*e.g.*, glycine, asparagine, glutamine, serine, threonine, tyrosine, cysteine), nonpolar side chains (*e.g.*, alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan), beta-branched side chains (*e.g.*, threonine, valine, isoleucine) and aromatic side chains (*e.g.*, tyrosine, phenylalanine, tryptophan, histidine). Thus, a predicted non-essential amino acid residue in the MOLX protein is replaced with another amino acid residue from the same side chain family. Alternatively, in another embodiment, mutations can be introduced randomly along all or part of an MOLX coding sequence, such as by saturation mutagenesis, and the resultant mutants can be screened for MOLX biological activity to identify mutants that retain activity. Following mutagenesis of SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, and 23, the encoded protein can be expressed by any recombinant technology known in the art and the activity of the protein can be determined.

The relatedness of amino acid families may also be determined based on side chain interactions. Substituted amino acids may be fully conserved "strong" residues or fully conserved "weak" residues. The "strong" group of conserved amino acid residues may be any one of the following groups: STA, NEQK, NHQK, NDEQ, QHRK, MILV, MILF, HY, FYW, wherein the single letter amino acid codes are grouped by those amino acids that may be substituted for each other. Likewise, the "weak" group of conserved residues may be any one of the following: CSA, ATV, SAG, STNK, STPA, SGND, SNDEQK, NDEQHK, NEQHRK, VLIM, HFY, wherein the letters within each group represent the single letter amino acid code.

In one embodiment, a mutant MOLX protein can be assayed for (*i*) the ability to form protein:protein interactions with other MOLX proteins, other cell-surface proteins, or biologically-active portions thereof, (*ii*) complex formation between a mutant MOLX protein and an MOLX ligand; or (*iii*) the ability of a mutant MOLX protein to bind to an intracellular target protein or biologically-active portion thereof; (*e.g.* avidin proteins).

In yet another embodiment, a mutant MOLX protein can be assayed for the ability to regulate a specific biological function (*e.g.*, regulation of insulin release).

Antisense Nucleic Acids

Another aspect of the invention pertains to isolated antisense nucleic acid molecules that are hybridizable to or complementary to the nucleic acid molecule comprising the nucleotide sequence of SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, and 23, or fragments, analogs or derivatives thereof. An "antisense" nucleic acid comprises a nucleotide sequence that is complementary to a "sense" nucleic acid encoding a protein (*e.g.*, complementary to the coding strand of a double-stranded cDNA molecule or complementary to an mRNA sequence). In specific aspects, antisense nucleic acid molecules are provided that comprise a sequence complementary to at least about 10, 25, 50, 100, 250 or 500 nucleotides or an entire MOLX coding strand, or to only a portion thereof. Nucleic acid molecules encoding fragments, homologs, derivatives and analogs of an MOLX protein of SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, and 24, or antisense nucleic acids complementary to an MOLX nucleic acid sequence of SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, and 23, are additionally provided.

In one embodiment, an antisense nucleic acid molecule is antisense to a "coding region" of the coding strand of a nucleotide sequence encoding an MOLX protein. The term "coding region" refers to the region of the nucleotide sequence comprising codons which are translated into amino acid residues. In another embodiment, the antisense nucleic acid molecule is antisense to a "noncoding region" of the coding strand of a nucleotide sequence encoding the MOLX protein. The term "noncoding region" refers to 5' and 3' sequences which flank the coding region that are not translated into amino acids (*i.e.*, also referred to as 5' and 3' untranslated regions).

Given the coding strand sequences encoding the MOLX protein disclosed herein, antisense nucleic acids of the invention can be designed according to the rules of Watson and Crick or Hoogsteen base pairing. The antisense nucleic acid molecule can be complementary to the entire coding region of MOLX mRNA, but more preferably is an oligonucleotide that is

antisense to only a portion of the coding or noncoding region of MOLX mRNA. For example, the antisense oligonucleotide can be complementary to the region surrounding the translation start site of MOLX mRNA. An antisense oligonucleotide can be, for example, about 5, 10, 15, 20, 25, 30, 35, 40, 45 or 50 nucleotides in length. An antisense nucleic acid of the invention can be constructed using chemical synthesis or enzymatic ligation reactions using procedures known in the art. For example, an antisense nucleic acid (*e.g.*, an antisense oligonucleotide) can be chemically synthesized using naturally-occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleic acids (*e.g.*, phosphorothioate derivatives and acridine substituted nucleotides can be used).

Examples of modified nucleotides that can be used to generate the antisense nucleic acid include: 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xanthine, 4-acetylcytosine, 5-(carboxyhydroxylmethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (v), wybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid (v), 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, (acp3)w, and 2,6-diaminopurine. Alternatively, the antisense nucleic acid can be produced biologically using an expression vector into which a nucleic acid has been subcloned in an antisense orientation (*i.e.*, RNA transcribed from the inserted nucleic acid will be of an antisense orientation to a target nucleic acid of interest, described further in the following subsection).

The antisense nucleic acid molecules of the invention are typically administered to a subject or generated *in situ* such that they hybridize with or bind to cellular mRNA and/or genomic DNA encoding an MOLX protein to thereby inhibit expression of the protein (*e.g.*, by inhibiting transcription and/or translation). The hybridization can be by conventional nucleotide complementarity to form a stable duplex, or, for example, in the case of an antisense nucleic acid molecule that binds to DNA duplexes, through specific interactions in the major groove of the double helix. An example of a route of administration of antisense nucleic acid molecules of the invention includes direct injection at a tissue site. Alternatively,

antisense nucleic acid molecules can be modified to target selected cells and then administered systemically. For example, for systemic administration, antisense molecules can be modified such that they specifically bind to receptors or antigens expressed on a selected cell surface (*e.g.*, by linking the antisense nucleic acid molecules to peptides or antibodies that bind to cell surface receptors or antigens). The antisense nucleic acid molecules can also be delivered to cells using the vectors described herein. To achieve sufficient nucleic acid molecules, vector constructs in which the antisense nucleic acid molecule is placed under the control of a strong pol II or pol III promoter are preferred.

In yet another embodiment, the antisense nucleic acid molecule of the invention is an α-anomeric nucleic acid molecule. An α-anomeric nucleic acid molecule forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual β-units, the strands run parallel to each other. See, *e.g.,* Gaultier, *et al.*, 1987. *Nucl. Acids Res.* **15**: 6625-6641. The antisense nucleic acid molecule can also comprise a 2'-o-methylribonucleotide (see, *e.g.,* Inoue, *et al.* 1987. *Nucl. Acids Res.* **15**: 6131-6148) or a chimeric RNA-DNA analogue (see, *e.g.,* Inoue, *et al.*, 1987. *FEBS Lett.* **215**: 327-330.

Ribozymes and PNA Moieties

Nucleic acid modifications include, by way of non-limiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject.

In one embodiment, an antisense nucleic acid of the invention is a ribozyme. Ribozymes are catalytic RNA molecules with ribonuclease activity that are capable of cleaving a single-stranded nucleic acid, such as an mRNA, to which they have a complementary region. Thus, ribozymes (*e.g.*, hammerhead ribozymes as described in Haselhoff and Gerlach 1988. *Nature* 334: 585-591) can be used to catalytically cleave MOLX mRNA transcripts to thereby inhibit translation of MOLX mRNA. A ribozyme having specificity for an MOLX-encoding nucleic acid can be designed based upon the nucleotide sequence of an MOLX cDNA disclosed herein (*i.e.*, SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, and 23). For example, a derivative of a *Tetrahymena* L-19 IVS RNA can be constructed in which the nucleotide sequence of the active site is complementary to the nucleotide sequence to be cleaved in an MOLX-encoding mRNA. See, *e.g.*, U.S. Patent 4,987,071 to Cech, *et al.* and U.S. Patent 5,116,742 to Cech, *et al.* MOLX mRNA can also be

used to select a catalytic RNA having a specific ribonuclease activity from a pool of RNA molecules. See, *e.g.*, Bartel *et al.*, (1993) *Science* 261:1411-1418.

Alternatively, MOLX gene expression can be inhibited by targeting nucleotide sequences complementary to the regulatory region of the MOLX nucleic acid (*e.g.*, the MOLX promoter and/or enhancers) to form triple helical structures that prevent transcription of the MOLX gene in target cells. *See, e.g.,* Helene, 1991. *Anticancer Drug Des.* 6: 569-84; Helene, *et al.* 1992. *Ann. N.Y. Acad. Sci.* 660: 27-36; Maher, 1992. *Bioassays* 14: 807-15.

In various embodiments, the MOLX nucleic acids can be modified at the base moiety, sugar moiety or phosphate backbone to improve, *e.g.*, the stability, hybridization, or solubility of the molecule. For example, the deoxyribose phosphate backbone of the nucleic acids can be modified to generate peptide nucleic acids. *See, e.g.,* Hyrup, *et al.*, 1996. *Bioorg Med Chem* 4: 5-23. As used herein, the terms "peptide nucleic acids" or "PNAs" refer to nucleic acid mimics (*e.g.*, DNA mimics) in which the deoxyribose phosphate backbone is replaced by a pseudopeptide backbone and only the four natural nucleobases are retained. The neutral backbone of PNAs has been shown to allow for specific hybridization to DNA and RNA under conditions of low ionic strength. The synthesis of PNA oligomers can be performed using standard solid phase peptide synthesis protocols as described in Hyrup, *et al.*, 1996. *supra*; Perry-O'Keefe, *et al.*, 1996. *Proc. Natl. Acad. Sci. USA* 93: 14670-14675.

PNAs of MOLX can be used in therapeutic and diagnostic applications. For example, PNAs can be used as antisense or antigene agents for sequence-specific modulation of gene expression by, *e.g.*, inducing transcription or translation arrest or inhibiting replication. PNAs of MOLX can also be used, for example, in the analysis of single base pair mutations in a gene (*e.g.*, PNA directed PCR clamping; as artificial restriction enzymes when used in combination with other enzymes, *e.g.*, S_1 nucleases (*see,* Hyrup, *et al.,* 1996.*supra*); or as probes or primers for DNA sequence and hybridization (*see,* Hyrup, *et al.*, 1996, *supra*; Perry-O'Keefe, *et al.,* 1996. *supra*).

In another embodiment, PNAs of MOLX can be modified, *e.g.*, to enhance their stability or cellular uptake, by attaching lipophilic or other helper groups to PNA, by the formation of PNA-DNA chimeras, or by the use of liposomes or other techniques of drug delivery known in the art. For example, PNA-DNA chimeras of MOLX can be generated that may combine the advantageous properties of PNA and DNA. Such chimeras allow DNA recognition enzymes (*e.g.*, RNase H and DNA polymerases) to interact with the DNA portion while the PNA portion would provide high binding affinity and specificity. PNA-DNA chimeras can be linked using linkers of appropriate lengths selected in terms of base stacking,

number of bonds between the nucleobases, and orientation (*see,* Hyrup, et al., 1996. *supra*). The synthesis of PNA-DNA chimeras can be performed as described in Hyrup, *et al.,* 1996. *supra* and Finn, *et al.*, 1996. *Nucl Acids Res* 24: 3357-3363. For example, a DNA chain can be synthesized on a solid support using standard phosphoramidite coupling chemistry, and modified nucleoside analogs, *e.g.*, 5'-(4-methoxytrityl)amino-5'-deoxy-thymidine phosphoramidite, can be used between the PNA and the 5' end of DNA. *See, e.g.,* Mag, *et al.*, 1989. *Nucl Acid Res* 17: 5973-5988. PNA monomers are then coupled in a stepwise manner to produce a chimeric molecule with a 5' PNA segment and a 3' DNA segment. *See, e.g.,* Finn, *et al.*, 1996. *supra*. Alternatively, chimeric molecules can be synthesized with a 5' DNA segment and a 3' PNA segment. *See, e.g.,* Petersen, *et al.*, 1975. *Bioorg. Med. Chem. Lett.* 5: 1119-11124.

In other embodiments, the oligonucleotide may include other appended groups such as peptides (*e.g.*, for targeting host cell receptors *in vivo*), or agents facilitating transport across the cell membrane (*see, e.g.*, Letsinger, *et al.*, 1989. *Proc. Natl. Acad. Sci. U.S.A.* 86: 6553-6556; Lemaitre, *et al.*, 1987. *Proc. Natl. Acad. Sci.* 84: 648-652; PCT Publication No. WO88/09810) or the blood-brain barrier (*see, e.g.*, PCT Publication No. WO 89/10134). In addition, oligonucleotides can be modified with hybridization triggered cleavage agents (*see, e.g.,* Krol, *et al.*, 1988. *BioTechniques* 6:958-976) or intercalating agents (*see, e.g.*, Zon, 1988. *Pharm. Res.* 5: 539-549). To this end, the oligonucleotide may be conjugated to another molecule, *e.g.*, a peptide, a hybridization triggered cross-linking agent, a transport agent, a hybridization-triggered cleavage agent, and the like.

MOLX Polypeptides

A polypeptide according to the invention includes a polypeptide including the amino acid sequence of MOLX polypeptides whose sequences are provided in SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, and 24. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residues shown in SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, and 24 while still encoding a protein that maintains its MOLX activities and physiological functions, or a functional fragment thereof.

In general, an MOLX variant that preserves MOLX-like function includes any variant in which residues at a particular position in the sequence have been substituted by other amino acids, and further include the possibility of inserting an additional residue or residues between two residues of the parent protein as well as the possibility of deleting one or more residues from the parent sequence. Any amino acid substitution, insertion, or deletion is encompassed

by the invention. In favorable circumstances, the substitution is a conservative substitution as defined above.

One aspect of the invention pertains to isolated MOLX proteins, and biologically-active portions thereof, or derivatives, fragments, analogs or homologs thereof. Also provided are polypeptide fragments suitable for use as immunogens to raise anti-MOLX antibodies. In one embodiment, native MOLX proteins can be isolated from cells or tissue sources by an appropriate purification scheme using standard protein purification techniques. In another embodiment, MOLX proteins are produced by recombinant DNA techniques. Alternative to recombinant expression, an MOLX protein or polypeptide can be synthesized chemically using standard peptide synthesis techniques.

An "isolated" or "purified" polypeptide or protein or biologically-active portion thereof is substantially free of cellular material or other contaminating proteins from the cell or tissue source from which the MOLX protein is derived, or substantially free from chemical precursors or other chemicals when chemically synthesized. The language "substantially free of cellular material" includes preparations of MOLX proteins in which the protein is separated from cellular components of the cells from which it is isolated or recombinantly-produced. In one embodiment, the language "substantially free of cellular material" includes preparations of MOLX proteins having less than about 30% (by dry weight) of non-MOLX proteins (also referred to herein as a "contaminating protein"), more preferably less than about 20% of non-MOLX proteins, still more preferably less than about 10% of non-MOLX proteins, and most preferably less than about 5% of non-MOLX proteins. When the MOLX protein or biologically-active portion thereof is recombinantly-produced, it is also preferably substantially free of culture medium, *i.e.*, culture medium represents less than about 20%, more preferably less than about 10%, and most preferably less than about 5% of the volume of the MOLX protein preparation.

The language "substantially free of chemical precursors or other chemicals" includes preparations of MOLX proteins in which the protein is separated from chemical precursors or other chemicals that are involved in the synthesis of the protein. In one embodiment, the language "substantially free of chemical precursors or other chemicals" includes preparations of MOLX proteins having less than about 30% (by dry weight) of chemical precursors or non-MOLX chemicals, more preferably less than about 20% chemical precursors or non-MOLX chemicals, still more preferably less than about 10% chemical precursors or non-MOLX chemicals, and most preferably less than about 5% chemical precursors or non-MOLX chemicals.

Biologically-active portions of MOLX proteins include peptides comprising amino acid sequences sufficiently homologous to or derived from the amino acid sequences of the MOLX proteins (*e.g.*, the amino acid sequence shown in SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, and 24) that include fewer amino acids than the full-length MOLX proteins, and exhibit at least one activity of an MOLX protein. Typically, biologically-active portions comprise a domain or motif with at least one activity of the MOLX protein. A biologically-active portion of an MOLX protein can be a polypeptide which is, for example, 10, 25, 50, 100 or more amino acid residues in length.

Moreover, other biologically-active portions, in which other regions of the protein are deleted, can be prepared by recombinant techniques and evaluated for one or more of the functional activities of a native MOLX protein.

In an embodiment, the MOLX protein has an amino acid sequence shown in SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, and 24. In other embodiments, the MOLX protein is substantially homologous to SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, and 24, and retains the functional activity of the protein of SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, and 24, yet differs in amino acid sequence due to natural allelic variation or mutagenesis, as described in detail, below. Accordingly, in another embodiment, the MOLX protein is a protein that comprises an amino acid sequence at least about 45% homologous to the amino acid sequence SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, and 24, and retains the functional activity of the MOLX proteins of SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, and 24.

Determining Homology Between Two or More Sequences

To determine the percent homology of two amino acid sequences or of two nucleic acids, the sequences are aligned for optimal comparison purposes (*e.g.*, gaps can be introduced in the sequence of a first amino acid or nucleic acid sequence for optimal alignment with a second amino or nucleic acid sequence). The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are homologous at that position (*i.e.*, as used herein amino acid or nucleic acid "homology" is equivalent to amino acid or nucleic acid "identity").

The nucleic acid sequence homology may be determined as the degree of identity between two sequences. The homology may be determined using computer programs known

in the art, such as GAP software provided in the GCG program package. *See,* Needleman and Wunsch, 1970. *J Mol Biol* 48: 443-453. Using GCG GAP software with the following settings for nucleic acid sequence comparison: GAP creation penalty of 5.0 and GAP extension penalty of 0.3, the coding region of the analogous nucleic acid sequences referred to above exhibits a degree of identity preferably of at least 70%, 75%, 80%, 85%, 90%, 95%, 98%, or 99%, with the CDS (encoding) part of the DNA sequence shown in SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, and 23.

The term "sequence identity" refers to the degree to which two polynucleotide or polypeptide sequences are identical on a residue-by-residue basis over a particular region of comparison. The term "percentage of sequence identity" is calculated by comparing two optimally aligned sequences over that region of comparison, determining the number of positions at which the identical nucleic acid base (*e.g.,* A, T, C, G, U, or I, in the case of nucleic acids) occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the region of comparison (*i.e.,* the window size), and multiplying the result by 100 to yield the percentage of sequence identity. The term "substantial identity" as used herein denotes a characteristic of a polynucleotide sequence, wherein the polynucleotide comprises a sequence that has at least 80 percent sequence identity, preferably at least 85 percent identity and often 90 to 95 percent sequence identity, more usually at least 99 percent sequence identity as compared to a reference sequence over a comparison region.

Chimeric and Fusion Proteins

The invention also provides MOLX chimeric or fusion proteins. As used herein, an MOLX "chimeric protein" or "fusion protein" comprises an MOLX polypeptide operatively-linked to a non-MOLX polypeptide. An "MOLX polypeptide" refers to a polypeptide having an amino acid sequence corresponding to an MOLX protein (SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, and 24), whereas a "non-MOLX polypeptide" refers to a polypeptide having an amino acid sequence corresponding to a protein that is not substantially homologous to the MOLX protein, *e.g.*, a protein that is different from the MOLX protein and that is derived from the same or a different organism. Within an MOLX fusion protein the MOLX polypeptide can correspond to all or a portion of an MOLX protein. In one embodiment, an MOLX fusion protein comprises at least one biologically-active portion of an MOLX protein. In another embodiment, an MOLX fusion protein comprises at least two biologically-active portions of an MOLX protein. In yet another embodiment, an MOLX fusion protein

comprises at least three biologically-active portions of an MOLX protein. Within the fusion protein, the term "operatively-linked" is intended to indicate that the MOLX polypeptide and the non-MOLX polypeptide are fused in-frame with one another. The non-MOLX polypeptide can be fused to the N-terminus or C-terminus of the MOLX polypeptide.

In one embodiment, the fusion protein is a GST-MOLX fusion protein in which the MOLX sequences are fused to the C-terminus of the GST (glutathione S-transferase) sequences. Such fusion proteins can facilitate the purification of recombinant MOLX polypeptides.

In another embodiment, the fusion protein is an MOLX protein containing a heterologous signal sequence at its N-terminus. In certain host cells (*e.g.*, mammalian host cells), expression and/or secretion of MOLX can be increased through use of a heterologous signal sequence.

In yet another embodiment, the fusion protein is an MOLX-immunoglobulin fusion protein in which the MOLX sequences are fused to sequences derived from a member of the immunoglobulin protein family. The MOLX-immunoglobulin fusion proteins of the invention can be incorporated into pharmaceutical compositions and administered to a subject to inhibit an interaction between an MOLX ligand and an MOLX protein on the surface of a cell, to thereby suppress MOLX-mediated signal transduction *in vivo*. The MOLX-immunoglobulin fusion proteins can be used to affect the bioavailability of an MOLX cognate ligand. Inhibition of the MOLX ligand/MOLX interaction may be useful therapeutically for both the treatment of proliferative and differentiative disorders, as well as modulating (*e.g.* promoting or inhibiting) cell survival. Moreover, the MOLX-immunoglobulin fusion proteins of the invention can be used as immunogens to produce anti-MOLX antibodies in a subject, to purify MOLX ligands, and in screening assays to identify molecules that inhibit the interaction of MOLX with an MOLX ligand.

An MOLX chimeric or fusion protein of the invention can be produced by standard recombinant DNA techniques. For example, DNA fragments coding for the different polypeptide sequences are ligated together in-frame in accordance with conventional techniques, *e.g.*, by employing blunt-ended or stagger-ended termini for ligation, restriction enzyme digestion to provide for appropriate termini, filling-in of cohesive ends as appropriate, alkaline phosphatase treatment to avoid undesirable joining, and enzymatic ligation. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor primers that give rise to complementary overhangs between two

consecutive gene fragments that can subsequently be annealed and reamplified to generate a chimeric gene sequence (*see, e.g.,* Ausubel, *et al.* (eds.) CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, 1992). Moreover, many expression vectors are commercially available that already encode a fusion moiety (*e.g.,* a GST polypeptide). An MOLX-encoding nucleic acid can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the MOLX protein.

MOLX Agonists and Antagonists

The invention also pertains to variants of the MOLX proteins that function as either MOLX agonists (*i.e.,* mimetics) or as MOLX antagonists. Variants of the MOLX protein can be generated by mutagenesis (*e.g.,* discrete point mutation or truncation of the MOLX protein). An agonist of the MOLX protein can retain substantially the same, or a subset of, the biological activities of the naturally occurring form of the MOLX protein. An antagonist of the MOLX protein can inhibit one or more of the activities of the naturally occurring form of the MOLX protein by, for example, competitively binding to a downstream or upstream member of a cellular signaling cascade which includes the MOLX protein. Thus, specific biological effects can be elicited by treatment with a variant of limited function. In one embodiment, treatment of a subject with a variant having a subset of the biological activities of the naturally occurring form of the protein has fewer side effects in a subject relative to treatment with the naturally occurring form of the MOLX proteins.

Variants of the MOLX proteins that function as either MOLX agonists (*i.e.,* mimetics) or as MOLX antagonists can be identified by screening combinatorial libraries of mutants (*e.g.,* truncation mutants) of the MOLX proteins for MOLX protein agonist or antagonist activity. In one embodiment, a variegated library of MOLX variants is generated by combinatorial mutagenesis at the nucleic acid level and is encoded by a variegated gene library. A variegated library of MOLX variants can be produced by, for example, enzymatically ligating a mixture of synthetic oligonucleotides into gene sequences such that a degenerate set of potential MOLX sequences is expressible as individual polypeptides, or alternatively, as a set of larger fusion proteins (*e.g.,* for phage display) containing the set of MOLX sequences therein. There are a variety of methods which can be used to produce libraries of potential MOLX variants from a degenerate oligonucleotide sequence. Chemical synthesis of a degenerate gene sequence can be performed in an automatic DNA synthesizer, and the synthetic gene then ligated into an appropriate expression vector. Use of a degenerate set of genes allows for the provision, in one mixture, of all of the sequences encoding the

desired set of potential MOLX sequences. Methods for synthesizing degenerate oligonucleotides are well-known within the art. *See, e.g.,* Narang, 1983. *Tetrahedron* 39: 3; Itakura, *et al.*, 1984. *Annu. Rev. Biochem.* 53: 323; Itakura, *et al.*, 1984. *Science* 198: 1056; Ike, *et al.*, 1983. *Nucl. Acids Res.* 11: 477.

Polypeptide Libraries

In addition, libraries of fragments of the MOLX protein coding sequences can be used to generate a variegated population of MOLX fragments for screening and subsequent selection of variants of an MOLX protein. In one embodiment, a library of coding sequence fragments can be generated by treating a double stranded PCR fragment of an MOLX coding sequence with a nuclease under conditions wherein nicking occurs only about once per molecule, denaturing the double stranded DNA, renaturing the DNA to form double-stranded DNA that can include sense/antisense pairs from different nicked products, removing single stranded portions from reformed duplexes by treatment with S_1 nuclease, and ligating the resulting fragment library into an expression vector. By this method, expression libraries can be derived which encodes N-terminal and internal fragments of various sizes of the MOLX proteins.

Various techniques are known in the art for screening gene products of combinatorial libraries made by point mutations or truncation, and for screening cDNA libraries for gene products having a selected property. Such techniques are adaptable for rapid screening of the gene libraries generated by the combinatorial mutagenesis of MOLX proteins. The most widely used techniques, which are amenable to high throughput analysis, for screening large gene libraries typically include cloning the gene library into replicable expression vectors, transforming appropriate cells with the resulting library of vectors, and expressing the combinatorial genes under conditions in which detection of a desired activity facilitates isolation of the vector encoding the gene whose product was detected. Recursive ensemble mutagenesis (REM), a new technique that enhances the frequency of functional mutants in the libraries, can be used in combination with the screening assays to identify MOLX variants. *See, e.g.,* Arkin and Yourvan, 1992. *Proc. Natl. Acad. Sci. USA* 89: 7811-7815; Delgrave, *et al.*, 1993. *Protein Engineering* 6:327-331.

Anti-MOLX Antibodies

The invention encompasses antibodies and antibody fragments, such as F_{ab} or $(F_{ab})_2$, that bind immunospecifically to any of the MOLX polypeptides of said invention.

An isolated MOLX protein, or a portion or fragment thereof, can be used as an immunogen to generate antibodies that bind to MOLX polypeptides using standard techniques for polyclonal and monoclonal antibody preparation. The full-length MOLX proteins can be used or, alternatively, the invention provides antigenic peptide fragments of MOLX proteins for use as immunogens. The antigenic MOLX peptides comprises at least 4 amino acid residues of the amino acid sequence shown in SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, and 24 and encompasses an epitope of MOLX such that an antibody raised against the peptide forms a specific immune complex with MOLX. Preferably, the antigenic peptide comprises at least 6, 8, 10, 15, 20, or 30 amino acid residues. Longer antigenic peptides are sometimes preferable over shorter antigenic peptides, depending on use and according to methods well known to someone skilled in the art.

In certain embodiments of the invention, at least one epitope encompassed by the antigenic peptide is a region of MOLX that is located on the surface of the protein (*e.g.*, a hydrophilic region). As a means for targeting antibody production, hydropathy plots showing regions of hydrophilicity and hydrophobicity may be generated by any method well known in the art, including, for example, the Kyte Doolittle or the Hopp Woods methods, either with or without Fourier transformation (*see, e.g.*, Hopp and Woods, 1981. *Proc. Nat. Acad. Sci. USA* 78: 3824-3828; Kyte and Doolittle, 1982. *J. Mol. Biol.* 157: 105-142, each incorporated herein by reference in their entirety).

As disclosed herein, MOLX protein sequences of SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, and 24, or derivatives, fragments, analogs or homologs thereof, may be utilized as immunogens in the generation of antibodies that immunospecifically-bind these protein components. The term "antibody" as used herein refers to immunoglobulin molecules and immunologically-active portions of immunoglobulin molecules, *i.e.*, molecules that contain an antigen binding site that specifically-binds (immunoreacts with) an antigen, such as MOLX. Such antibodies include, but are not limited to, polyclonal, monoclonal, chimeric, single chain, F_{ab} and $F_{(ab')2}$ fragments, and an F_{ab} expression library. In a specific embodiment, antibodies to human MOLX proteins are disclosed. Various procedures known within the art may be used for the production of polyclonal or monoclonal antibodies to an MOLX protein sequence of SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, and 24, or a derivative, fragment, analog or homolog thereof. Some of these proteins are discussed below.

For the production of polyclonal antibodies, various suitable host animals (*e.g.*, rabbit, goat, mouse or other mammal) may be immunized by injection with the native protein, or a synthetic variant thereof, or a derivative of the foregoing. An appropriate immunogenic

preparation can contain, for example, recombinantly-expressed MOLX protein or a chemically-synthesized MOLX polypeptide. The preparation can further include an adjuvant. Various adjuvants used to increase the immunological response include, but are not limited to, Freund's (complete and incomplete), mineral gels (*e.g.*, aluminum hydroxide), surface active substances (*e.g.*, lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, dinitrophenol, etc.), human adjuvants such as *Bacille Calmette-Guerin* and *Corynebacterium parvum*, or similar immunostimulatory agents. If desired, the antibody molecules directed against MOLX can be isolated from the mammal (*e.g.*, from the blood) and further purified by well known techniques, such as protein A chromatography to obtain the IgG fraction.

The term "monoclonal antibody" or "monoclonal antibody composition", as used herein, refers to a population of antibody molecules that contain only one species of an antigen binding site capable of immunoreacting with a particular epitope of MOLX. A monoclonal antibody composition thus typically displays a single binding affinity for a particular MOLX protein with which it immunoreacts. For preparation of monoclonal antibodies directed towards a particular MOLX protein, or derivatives, fragments, analogs or homologs thereof, any technique that provides for the production of antibody molecules by continuous cell line culture may be utilized. Such techniques include, but are not limited to, the hybridoma technique (*see, e.g.*, Kohler & Milstein, 1975. *Nature* 256: 495-497); the trioma technique; the human B-cell hybridoma technique (*see, e.g.*, Kozbor, *et al.*, 1983. *Immunol. Today* 4: 72) and the EBV hybridoma technique to produce human monoclonal antibodies (*see, e.g.*, Cole, *et al.*, 1985. In: MONOCLONAL ANTIBODIES AND CANCER THERAPY, Alan R. Liss, Inc., pp. 77-96). Human monoclonal antibodies may be utilized in the practice of the invention and may be produced by using human hybridomas (*see, e.g.*, Cote, *et al.*, 1983. *Proc Natl Acad Sci USA* 80: 2026-2030) or by transforming human B-cells with Epstein Barr Virus *in vitro* (*see, e.g.*, Cole, *et al.*, 1985. In: MONOCLONAL ANTIBODIES AND CANCER THERAPY, Alan R. Liss, Inc., pp. 77-96). Each of the above citations is incorporated herein by reference in their entirety.

According to the invention, techniques can be adapted for the production of single-chain antibodies specific to an MOLX protein (*see, e.g.*, U.S. Patent No. 4,946,778). In addition, methods can be adapted for the construction of F_{ab} expression libraries (*see, e.g.*, Huse, *et al.*, 1989. *Science* 246: 1275-1281) to allow rapid and effective identification of monoclonal F_{ab} fragments with the desired specificity for an MOLX protein or derivatives, fragments, analogs or homologs thereof. Non-human antibodies can be "humanized" by techniques well known in the art. *See, e.g.*, U.S. Patent No. 5,225,539. Antibody fragments that contain the idiotypes to an MOLX protein may be produced by techniques known in the

art including, but not limited to: (*i*) an $F_{(ab')2}$ fragment produced by pepsin digestion of an antibody molecule; (*ii*) an F_{ab} fragment generated by reducing the disulfide bridges of an $F_{(ab')2}$ fragment; (*iii*) an F_{ab} fragment generated by the treatment of the antibody molecule with papain and a reducing agent; and (*iv*) F_v fragments.

Additionally, recombinant anti-MOLX antibodies, such as chimeric and humanized monoclonal antibodies, comprising both human and non-human portions, which can be made using standard recombinant DNA techniques, are within the scope of the invention. Such chimeric and humanized monoclonal antibodies can be produced by recombinant DNA techniques known in the art, for example using methods described in International Application No. PCT/US86/02269; European Patent Application No. 184,187; European Patent Application No. 171,496; European Patent Application No. 173,494; PCT International Publication No. WO 86/01533; U.S. Patent No. 4,816,567; U.S. Pat. No. 5,225,539; European Patent Application No. 125,023; Better, *et al.*, 1988. *Science* 240: 1041-1043; Liu, *et al.*, 1987. *Proc. Natl. Acad. Sci. USA* 84: 3439-3443; Liu, *et al.*, 1987. *J. Immunol.* 139: 3521-3526; Sun, *et al.*, 1987. *Proc. Natl. Acad. Sci. USA* 84: 214-218; Nishimura, *et al.*, 1987. *Cancer Res.* 47: 999-1005; Wood, *et al.*, 1985. *Nature* 314 :446-449; Shaw, *et al.*, 1988. *J. Natl. Cancer Inst.* 80: 1553-1559); Morrison(1985) *Science* 229:1202-1207; Oi, *et al.* (1986) *BioTechniques* 4:214; Jones, *et al.*, 1986. *Nature* 321: 552-525; Verhoeyan, *et al.*, 1988. *Science* 239: 1534; and Beidler, *et al.*, 1988. *J. Immunol.* 141: 4053-4060. Each of the above citations are incorporated herein by reference in their entirety.

In one embodiment, methods for the screening of antibodies that possess the desired specificity include, but are not limited to, enzyme-linked immunosorbent assay (ELISA) and other immunologically-mediated techniques known within the art. In a specific embodiment, selection of antibodies that are specific to a particular domain of an MOLX protein is facilitated by generation of hybridomas that bind to the fragment of an MOLX protein possessing such a domain. Thus, antibodies that are specific for a desired domain within an MOLX protein, or derivatives, fragments, analogs or homologs thereof, are also provided herein.

Anti-MOLX antibodies may be used in methods known within the art relating to the localization and/or quantitation of an MOLX protein (*e.g.*, for use in measuring levels of the MOLX protein within appropriate physiological samples, for use in diagnostic methods, for use in imaging the protein, and the like). In a given embodiment, antibodies for MOLX proteins, or derivatives, fragments, analogs or homologs thereof, that contain the antibody

derived binding domain, are utilized as pharmacologically-active compounds (hereinafter "Therapeutics").

An anti-MOLX antibody (*e.g.*, monoclonal antibody) can be used to isolate an MOLX polypeptide by standard techniques, such as affinity chromatography or immunoprecipitation. An anti-MOLX antibody can facilitate the purification of natural MOLX polypeptide from cells and of recombinantly-produced MOLX polypeptide expressed in host cells. Moreover, an anti-MOLX antibody can be used to detect MOLX protein (*e.g.*, in a cellular lysate or cell supernatant) in order to evaluate the abundance and pattern of expression of the MOLX protein. Anti-MOLX antibodies can be used diagnostically to monitor protein levels in tissue as part of a clinical testing procedure, *e.g.*, to, for example, determine the efficacy of a given treatment regimen. Detection can be facilitated by coupling (*i.e.*, physically linking) the antibody to a detectable substance. Examples of detectable substances include various enzymes, prosthetic groups, fluorescent materials, luminescent materials, bioluminescent materials, and radioactive materials. Examples of suitable enzymes include horseradish peroxidase, alkaline phosphatase, β-galactosidase, or acetylcholinesterase; examples of suitable prosthetic group complexes include streptavidin/biotin and avidin/biotin; examples of suitable fluorescent materials include umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or phycoerythrin; an example of a luminescent material includes luminol; examples of bioluminescent materials include luciferase, luciferin, and aequorin, and examples of suitable radioactive material include ^{125}I, ^{131}I, ^{35}S or ^{3}H.

MOLX Recombinant Expression Vectors and Host Cells

Another aspect of the invention pertains to vectors, preferably expression vectors, containing a nucleic acid encoding an MOLX protein, or derivatives, fragments, analogs or homologs thereof. As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of vector is a "plasmid", which refers to a circular double stranded DNA loop into which additional DNA segments can be ligated. Another type of vector is a viral vector, wherein additional DNA segments can be ligated into the viral genome. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (*e.g.*, bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Other vectors (*e.g.*, non-episomal mammalian vectors) are integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome.

Moreover, certain vectors are capable of directing the expression of genes to which they are operatively-linked. Such vectors are referred to herein as "expression vectors". In general, expression vectors of utility in recombinant DNA techniques are often in the form of plasmids. In the present specification, "plasmid" and "vector" can be used interchangeably as the plasmid is the most commonly used form of vector. However, the invention is intended to include such other forms of expression vectors, such as viral vectors (*e.g.*, replication defective retroviruses, adenoviruses and adeno-associated viruses), which serve equivalent functions.

The recombinant expression vectors of the invention comprise a nucleic acid of the invention in a form suitable for expression of the nucleic acid in a host cell, which means that the recombinant expression vectors include one or more regulatory sequences, selected on the basis of the host cells to be used for expression, that is operatively-linked to the nucleic acid sequence to be expressed. Within a recombinant expression vector, "operably-linked" is intended to mean that the nucleotide sequence of interest is linked to the regulatory sequence(s) in a manner that allows for expression of the nucleotide sequence (*e.g.*, in an *in vitro* transcription/translation system or in a host cell when the vector is introduced into the host cell).

The term "regulatory sequence" is intended to includes promoters, enhancers and other expression control elements (*e.g.*, polyadenylation signals). Such regulatory sequences are described, for example, in Goeddel, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990). Regulatory sequences include those that direct constitutive expression of a nucleotide sequence in many types of host cell and those that direct expression of the nucleotide sequence only in certain host cells (*e.g.*, tissue-specific regulatory sequences). It will be appreciated by those skilled in the art that the design of the expression vector can depend on such factors as the choice of the host cell to be transformed, the level of expression of protein desired, etc. The expression vectors of the invention can be introduced into host cells to thereby produce proteins or peptides, including fusion proteins or peptides, encoded by nucleic acids as described herein (*e.g.*, MOLX proteins, mutant forms of MOLX proteins, fusion proteins, etc.).

The recombinant expression vectors of the invention can be designed for expression of MOLX proteins in prokaryotic or eukaryotic cells. For example, MOLX proteins can be expressed in bacterial cells such as *Escherichia coli*, insect cells (using baculovirus expression vectors) yeast cells or mammalian cells. Suitable host cells are discussed further in Goeddel, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San

Diego, Calif. (1990). Alternatively, the recombinant expression vector can be transcribed and translated *in vitro*, for example using T7 promoter regulatory sequences and T7 polymerase.

Expression of proteins in prokaryotes is most often carried out in *Escherichia coli* with vectors containing constitutive or inducible promoters directing the expression of either fusion or non-fusion proteins. Fusion vectors add a number of amino acids to a protein encoded therein, usually to the amino terminus of the recombinant protein. Such fusion vectors typically serve three purposes: (*i*) to increase expression of recombinant protein; (*ii*) to increase the solubility of the recombinant protein; and (*iii*) to aid in the purification of the recombinant protein by acting as a ligand in affinity purification. Often, in fusion expression vectors, a proteolytic cleavage site is introduced at the junction of the fusion moiety and the recombinant protein to enable separation of the recombinant protein from the fusion moiety subsequent to purification of the fusion protein. Such enzymes, and their cognate recognition sequences, include Factor Xa, thrombin and enterokinase. Typical fusion expression vectors include pGEX (Pharmacia Biotech Inc; Smith and Johnson, 1988. *Gene* 67: 31-40), pMAL (New England Biolabs, Beverly, Mass.) and pRIT5 (Pharmacia, Piscataway, N.J.) that fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein.

Examples of suitable inducible non-fusion *E. coli* expression vectors include pTrc (Amrann *et al.*, (1988) *Gene* 69:301-315) and pET 11d (Studier *et al.*, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990) 60-89).

One strategy to maximize recombinant protein expression in *E. coli* is to express the protein in a host bacteria with an impaired capacity to proteolytically cleave the recombinant protein. *See, e.g.,* Gottesman, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990) 119-128. Another strategy is to alter the nucleic acid sequence of the nucleic acid to be inserted into an expression vector so that the individual codons for each amino acid are those preferentially utilized in *E. coli* (*see, e.g.,* Wada, *et al.*, 1992. *Nucl. Acids Res.* 20: 2111-2118). Such alteration of nucleic acid sequences of the invention can be carried out by standard DNA synthesis techniques.

In another embodiment, the MOLX expression vector is a yeast expression vector. Examples of vectors for expression in yeast *Saccharomyces cerivisae* include pYepSec1 (Baldari, *et al.*, 1987. *EMBO J.* 6: 229-234), pMFa (Kurjan and Herskowitz, 1982. *Cell* 30: 933-943), pJRY88 (Schultz *et al.*, 1987. *Gene* 54: 113-123), pYES2 (Invitrogen Corporation, San Diego, Calif.), and picZ (InVitrogen Corp, San Diego, Calif.).

Alternatively, MOLX can be expressed in insect cells using baculovirus expression vectors. Baculovirus vectors available for expression of proteins in cultured insect cells (*e.g.*, SF9 cells) include the pAc series (Smith, *et al.*, 1983. *Mol. Cell. Biol.* 3: 2156-2165) and the pVL series (Lucklow and Summers, 1989. *Virology* 170: 31-39).

In yet another embodiment, a nucleic acid of the invention is expressed in mammalian cells using a mammalian expression vector. Examples of mammalian expression vectors include pCDM8 (Seed, 1987. *Nature* 329: 840) and pMT2PC (Kaufman, *et al.*, 1987. *EMBO J.* 6: 187-195). When used in mammalian cells, the expression vector's control functions are often provided by viral regulatory elements. For example, commonly used promoters are derived from polyoma, adenovirus 2, cytomegalovirus, and simian virus 40. For other suitable expression systems for both prokaryotic and eukaryotic cells see, *e.g.*, Chapters 16 and 17 of Sambrook, *et al.*, MOLECULAR CLONING: A LABORATORY MANUAL. 2nd ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989.

In another embodiment, the recombinant mammalian expression vector is capable of directing expression of the nucleic acid preferentially in a particular cell type (*e.g.*, tissue-specific regulatory elements are used to express the nucleic acid). Tissue-specific regulatory elements are known in the art. Non-limiting examples of suitable tissue-specific promoters include the albumin promoter (liver-specific; Pinkert, *et al.*, 1987. *Genes Dev.* 1: 268-277), lymphoid-specific promoters (Calame and Eaton, 1988. *Adv. Immunol.* 43: 235-275), in particular promoters of T cell receptors (Winoto and Baltimore, 1989. *EMBO J.* 8: 729-733) and immunoglobulins (Banerji, *et al.*, 1983. *Cell* 33: 729-740; Queen and Baltimore, 1983. *Cell* 33: 741-748), neuron-specific promoters (*e.g.*, the neurofilament promoter; Byrne and Ruddle, 1989. *Proc. Natl. Acad. Sci. USA* 86: 5473-5477), pancreas-specific promoters (Edlund, *et al.*, 1985. *Science* 230: 912-916), and mammary gland-specific promoters (*e.g.*, milk whey promoter; U.S. Pat. No. 4,873,316 and European Application Publication No. 264,166). Developmentally-regulated promoters are also encompassed, *e.g.*, the murine hox promoters (Kessel and Gruss, 1990. *Science* 249: 374-379) and the α-fetoprotein promoter (Campes and Tilghman, 1989. *Genes Dev.* 3: 537-546).

The invention further provides a recombinant expression vector comprising a DNA molecule of the invention cloned into the expression vector in an antisense orientation. That is, the DNA molecule is operatively-linked to a regulatory sequence in a manner that allows for expression (by transcription of the DNA molecule) of an RNA molecule that is antisense to MOLX mRNA. Regulatory sequences operatively linked to a nucleic acid cloned in the antisense orientation can be chosen that direct the continuous expression of the antisense RNA

molecule in a variety of cell types, for instance viral promoters and/or enhancers, or regulatory sequences can be chosen that direct constitutive, tissue specific or cell type specific expression of antisense RNA. The antisense expression vector can be in the form of a recombinant plasmid, phagemid or attenuated virus in which antisense nucleic acids are produced under the control of a high efficiency regulatory region, the activity of which can be determined by the cell type into which the vector is introduced. For a discussion of the regulation of gene expression using antisense genes *see, e.g.,* Weintraub, *et al.*, "Antisense RNA as a molecular tool for genetic analysis," *Reviews-Trends in Genetics*, Vol. 1(1) 1986.

Another aspect of the invention pertains to host cells into which a recombinant expression vector of the invention has been introduced. The terms "host cell" and "recombinant host cell" are used interchangeably herein. It is understood that such terms refer not only to the particular subject cell but also to the progeny or potential progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein.

A host cell can be any prokaryotic or eukaryotic cell. For example, MOLX protein can be expressed in bacterial cells such as *E. coli*, insect cells, yeast or mammalian cells (such as Chinese hamster ovary cells (CHO) or COS cells). Other suitable host cells are known to those skilled in the art.

Vector DNA can be introduced into prokaryotic or eukaryotic cells via conventional transformation or transfection techniques. As used herein, the terms "transformation" and "transfection" are intended to refer to a variety of art-recognized techniques for introducing foreign nucleic acid (*e.g.*, DNA) into a host cell, including calcium phosphate or calcium chloride co-precipitation, DEAE-dextran-mediated transfection, lipofection, or electroporation. Suitable methods for transforming or transfecting host cells can be found in Sambrook, *et al.* (MOLECULAR CLONING: A LABORATORY MANUAL. 2nd ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989), and other laboratory manuals.

For stable transfection of mammalian cells, it is known that, depending upon the expression vector and transfection technique used, only a small fraction of cells may integrate the foreign DNA into their genome. In order to identify and select these integrants, a gene that encodes a selectable marker (*e.g.*, resistance to antibiotics) is generally introduced into the host cells along with the gene of interest. Various selectable markers include those that confer resistance to drugs, such as G418, hygromycin and methotrexate. Nucleic acid encoding a

selectable marker can be introduced into a host cell on the same vector as that encoding MOLX or can be introduced on a separate vector. Cells stably transfected with the introduced nucleic acid can be identified by drug selection (*e.g.*, cells that have incorporated the selectable marker gene will survive, while the other cells die).

A host cell of the invention, such as a prokaryotic or eukaryotic host cell in culture, can be used to produce (*i.e.*, express) MOLX protein. Accordingly, the invention further provides methods for producing MOLX protein using the host cells of the invention. In one embodiment, the method comprises culturing the host cell of invention (into which a recombinant expression vector encoding MOLX protein has been introduced) in a suitable medium such that MOLX protein is produced. In another embodiment, the method further comprises isolating MOLX protein from the medium or the host cell.

Transgenic MOLX Animals

The host cells of the invention can also be used to produce non-human transgenic animals. For example, in one embodiment, a host cell of the invention is a fertilized oocyte or an embryonic stem cell into which MOLX protein-coding sequences have been introduced. Such host cells can then be used to create non-human transgenic animals in which exogenous MOLX sequences have been introduced into their genome or homologous recombinant animals in which endogenous MOLX sequences have been altered. Such animals are useful for studying the function and/or activity of MOLX protein and for identifying and/or evaluating modulators of MOLX protein activity. As used herein, a "transgenic animal" is a non-human animal, preferably a mammal, more preferably a rodent such as a rat or mouse, in which one or more of the cells of the animal includes a transgene. Other examples of transgenic animals include non-human primates, sheep, dogs, cows, goats, chickens, amphibians, etc. A transgene is exogenous DNA that is integrated into the genome of a cell from which a transgenic animal develops and that remains in the genome of the mature animal, thereby directing the expression of an encoded gene product in one or more cell types or tissues of the transgenic animal. As used herein, a "homologous recombinant animal" is a non-human animal, preferably a mammal, more preferably a mouse, in which an endogenous MOLX gene has been altered by homologous recombination between the endogenous gene and an exogenous DNA molecule introduced into a cell of the animal, *e.g.*, an embryonic cell of the animal, prior to development of the animal.

A transgenic animal of the invention can be created by introducing MOLX-encoding nucleic acid into the male pronuclei of a fertilized oocyte (*e.g.*, by microinjection, retroviral

infection) and allowing the oocyte to develop in a pseudopregnant female foster animal. The human MOLX cDNA sequences of SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, and 23 can be introduced as a transgene into the genome of a non-human animal. Alternatively, a non-human homologue of the human MOLX gene, such as a mouse MOLX gene, can be isolated based on hybridization to the human MOLX cDNA (described further *supra*) and used as a transgene. Intronic sequences and polyadenylation signals can also be included in the transgene to increase the efficiency of expression of the transgene. A tissue-specific regulatory sequence(s) can be operably-linked to the MOLX transgene to direct expression of MOLX protein to particular cells. Methods for generating transgenic animals via embryo manipulation and microinjection, particularly animals such as mice, have become conventional in the art and are described, for example, in U.S. Patent Nos. 4,736,866; 4,870,009; and 4,873,191; and Hogan, 1986. In: MANIPULATING THE MOUSE EMBRYO, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. Similar methods are used for production of other transgenic animals. A transgenic founder animal can be identified based upon the presence of the MOLX transgene in its genome and/or expression of MOLX mRNA in tissues or cells of the animals. A transgenic founder animal can then be used to breed additional animals carrying the transgene. Moreover, transgenic animals carrying a transgene-encoding MOLX protein can further be bred to other transgenic animals carrying other transgenes.

To create a homologous recombinant animal, a vector is prepared which contains at least a portion of an MOLX gene into which a deletion, addition or substitution has been introduced to thereby alter, *e.g.*, functionally disrupt, the MOLX gene. The MOLX gene can be a human gene (*e.g.*, the cDNA of SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, and 23), but more preferably, is a non-human homologue of a human MOLX gene. For example, a mouse homologue of human MOLX gene of SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, and 23 can be used to construct a homologous recombination vector suitable for altering an endogenous MOLX gene in the mouse genome. In one embodiment, the vector is designed such that, upon homologous recombination, the endogenous MOLX gene is functionally disrupted (*i.e.*, no longer encodes a functional protein; also referred to as a "knock out" vector).

Alternatively, the vector can be designed such that, upon homologous recombination, the endogenous MOLX gene is mutated or otherwise altered but still encodes functional protein (*e.g.*, the upstream regulatory region can be altered to thereby alter the expression of the endogenous MOLX protein). In the homologous recombination vector, the altered portion

of the MOLX gene is flanked at its 5'- and 3'-termini by additional nucleic acid of the MOLX gene to allow for homologous recombination to occur between the exogenous MOLX gene carried by the vector and an endogenous MOLX gene in an embryonic stem cell. The additional flanking MOLX nucleic acid is of sufficient length for successful homologous recombination with the endogenous gene. Typically, several kilobases of flanking DNA (both at the 5'- and 3'-termini) are included in the vector. *See, e.g.,* Thomas, *et al.*, 1987. *Cell* 51: 503 for a description of homologous recombination vectors. The vector is ten introduced into an embryonic stem cell line (*e.g.*, by electroporation) and cells in which the introduced MOLX gene has homologously-recombined with the endogenous MOLX gene are selected. *See, e.g.,* Li, *et al.*, 1992. *Cell* 69: 915.

The selected cells are then injected into a blastocyst of an animal (*e.g.*, a mouse) to form aggregation chimeras. *See, e.g.*, Bradley, 1987. In: TERATOCARCINOMAS AND EMBRYONIC STEM CELLS: A PRACTICAL APPROACH, Robertson, ed. IRL, Oxford, pp. 113-152. A chimeric embryo can then be implanted into a suitable pseudopregnant female foster animal and the embryo brought to term. Progeny harboring the homologously-recombined DNA in their germ cells can be used to breed animals in which all cells of the animal contain the homologously-recombined DNA by germline transmission of the transgene. Methods for constructing homologous recombination vectors and homologous recombinant animals are described further in Bradley, 1991. *Curr. Opin. Biotechnol.* 2: 823-829; PCT International Publication Nos.: WO 90/11354; WO 91/01140; WO 92/0968; and WO 93/04169.

In another embodiment, transgenic non-humans animals can be produced that contain selected systems that allow for regulated expression of the transgene. One example of such a system is the cre/loxP recombinase system of bacteriophage P1. For a description of the cre/loxP recombinase system, *See, e.g.*, Lakso, *et al.*, 1992. *Proc. Natl. Acad. Sci. USA* 89: 6232-6236. Another example of a recombinase system is the FLP recombinase system of *Saccharomyces cerevisiae. See,* O'Gorman, *et al.*, 1991. *Science* 251:1351-1355. If a cre/loxP recombinase system is used to regulate expression of the transgene, animals containing transgenes encoding both the Cre recombinase and a selected protein are required. Such animals can be provided through the construction of "double" transgenic animals, *e.g.*, by mating two transgenic animals, one containing a transgene encoding a selected protein and the other containing a transgene encoding a recombinase.

Clones of the non-human transgenic animals described herein can also be produced according to the methods described in Wilmut, *et al.*, 1997. *Nature* 385: 810-813. In brief, a cell (*e.g.*, a somatic cell) from the transgenic animal can be isolated and induced to exit the

growth cycle and enter G_0 phase. The quiescent cell can then be fused, *e.g.*, through the use of electrical pulses, to an enucleated oocyte from an animal of the same species from which the quiescent cell is isolated. The reconstructed oocyte is then cultured such that it develops to morula or blastocyte and then transferred to pseudopregnant female foster animal. The offspring borne of this female foster animal will be a clone of the animal from which the cell (*e.g.*, the somatic cell) is isolated.

Pharmaceutical Compositions

The MOLX nucleic acid molecules, MOLX proteins, and anti-MOLX antibodies (also referred to herein as "active compounds") of the invention, and derivatives, fragments, analogs and homologs thereof, can be incorporated into pharmaceutical compositions suitable for administration. Such compositions typically comprise the nucleic acid molecule, protein, or antibody and a pharmaceutically acceptable carrier. As used herein, "pharmaceutically acceptable carrier" is intended to include any and all solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents, and the like, compatible with pharmaceutical administration. Suitable carriers are described in the most recent edition of Remington's Pharmaceutical Sciences, a standard reference text in the field, which is incorporated herein by reference. Preferred examples of such carriers or diluents include, but are not limited to, water, saline, finger's solutions, dextrose solution, and 5% human serum albumin. Liposomes and non-aqueous vehicles such as fixed oils may also be used. The use of such media and agents for pharmaceutically active substances is well known in the art. Except insofar as any conventional media or agent is incompatible with the active compound, use thereof in the compositions is contemplated. Supplementary active compounds can also be incorporated into the compositions.

A pharmaceutical composition of the invention is formulated to be compatible with its intended route of administration. Examples of routes of administration include parenteral, *e.g.*, intravenous, intradermal, subcutaneous, oral (*e.g.*, inhalation), transdermal (*i.e.*, topical), transmucosal, and rectal administration. Solutions or suspensions used for parenteral, intradermal, or subcutaneous application can include the following components: a sterile diluent such as water for injection, saline solution, fixed oils, polyethylene glycols, glycerine, propylene glycol or other synthetic solvents; antibacterial agents such as benzyl alcohol or methyl parabens; antioxidants such as ascorbic acid or sodium bisulfite; chelating agents such as ethylenediaminetetraacetic acid (EDTA); buffers such as acetates, citrates or phosphates, and agents for the adjustment of tonicity such as sodium chloride or dextrose. The pH can be

adjusted with acids or bases, such as hydrochloric acid or sodium hydroxide. The parenteral preparation can be enclosed in ampoules, disposable syringes or multiple dose vials made of glass or plastic.

Pharmaceutical compositions suitable for injectable use include sterile aqueous solutions (where water soluble) or dispersions and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersion. For intravenous administration, suitable carriers include physiological saline, bacteriostatic water, Cremophor EL™ (BASF, Parsippany, N.J.) or phosphate buffered saline (PBS). In all cases, the composition must be sterile and should be fluid to the extent that easy syringeability exists. It must be stable under the conditions of manufacture and storage and must be preserved against the contaminating action of microorganisms such as bacteria and fungi. The carrier can be a solvent or dispersion medium containing, for example, water, ethanol, polyol (for example, glycerol, propylene glycol, and liquid polyethylene glycol, and the like), and suitable mixtures thereof. The proper fluidity can be maintained, for example, by the use of a coating such as lecithin, by the maintenance of the required particle size in the case of dispersion and by the use of surfactants. Prevention of the action of microorganisms can be achieved by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, ascorbic acid, thimerosal, and the like. In many cases, it will be preferable to include isotonic agents, for example, sugars, polyalcohols such as manitol, sorbitol, sodium chloride in the composition. Prolonged absorption of the injectable compositions can be brought about by including in the composition an agent which delays absorption, for example, aluminum monostearate and gelatin.

Sterile injectable solutions can be prepared by incorporating the active compound (*e.g.*, an MOLX protein or anti-MOLX antibody) in the required amount in an appropriate solvent with one or a combination of ingredients enumerated above, as required, followed by filtered sterilization. Generally, dispersions are prepared by incorporating the active compound into a sterile vehicle that contains a basic dispersion medium and the required other ingredients from those enumerated above. In the case of sterile powders for the preparation of sterile injectable solutions, methods of preparation are vacuum drying and freeze-drying that yields a powder of the active ingredient plus any additional desired ingredient from a previously sterile-filtered solution thereof.

Oral compositions generally include an inert diluent or an edible carrier. They can be enclosed in gelatin capsules or compressed into tablets. For the purpose of oral therapeutic administration, the active compound can be incorporated with excipients and used in the form

of tablets, troches, or capsules. Oral compositions can also be prepared using a fluid carrier for use as a mouthwash, wherein the compound in the fluid carrier is applied orally and swished and expectorated or swallowed. Pharmaceutically compatible binding agents, and/or adjuvant materials can be included as part of the composition. The tablets, pills, capsules, troches and the like can contain any of the following ingredients, or compounds of a similar nature: a binder such as microcrystalline cellulose, gum tragacanth or gelatin; an excipient such as starch or lactose, a disintegrating agent such as alginic acid, Primogel, or corn starch; a lubricant such as magnesium stearate or Sterotes; a glidant such as colloidal silicon dioxide; a sweetening agent such as sucrose or saccharin; or a flavoring agent such as peppermint, methyl salicylate, or orange flavoring.

For administration by inhalation, the compounds are delivered in the form of an aerosol spray from pressured container or dispenser which contains a suitable propellant, *e.g.*, a gas such as carbon dioxide, or a nebulizer.

Systemic administration can also be by transmucosal or transdermal means. For transmucosal or transdermal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art, and include, for example, for transmucosal administration, detergents, bile salts, and fusidic acid derivatives. Transmucosal administration can be accomplished through the use of nasal sprays or suppositories. For transdermal administration, the active compounds are formulated into ointments, salves, gels, or creams as generally known in the art.

The compounds can also be prepared in the form of suppositories (*e.g.*, with conventional suppository bases such as cocoa butter and other glycerides) or retention enemas for rectal delivery.

In one embodiment, the active compounds are prepared with carriers that will protect the compound against rapid elimination from the body, such as a controlled release formulation, including implants and microencapsulated delivery systems. Biodegradable, biocompatible polymers can be used, such as ethylene vinyl acetate, polyanhydrides, polyglycolic acid, collagen, polyorthoesters, and polylactic acid. Methods for preparation of such formulations will be apparent to those skilled in the art. The materials can also be obtained commercially from Alza Corporation and Nova Pharmaceuticals, Inc. Liposomal suspensions (including liposomes targeted to infected cells with monoclonal antibodies to viral antigens) can also be used as pharmaceutically acceptable carriers. These can be prepared according to methods known to those skilled in the art, for example, as described in U.S. Patent No. 4,522,811.

It is especially advantageous to formulate oral or parenteral compositions in dosage unit form for ease of administration and uniformity of dosage. Dosage unit form as used herein refers to physically discrete units suited as unitary dosages for the subject to be treated; each unit containing a predetermined quantity of active compound calculated to produce the desired therapeutic effect in association with the required pharmaceutical carrier. The specification for the dosage unit forms of the invention are dictated by and directly dependent on the unique characteristics of the active compound and the particular therapeutic effect to be achieved, and the limitations inherent in the art of compounding such an active compound for the treatment of individuals.

The nucleic acid molecules of the invention can be inserted into vectors and used as gene therapy vectors. Gene therapy vectors can be delivered to a subject by, for example, intravenous injection, local administration (*see, e.g.,* U.S. Patent No. 5,328,470) or by stereotactic injection (*see, e.g.*, Chen, *et al.*, 1994. *Proc. Natl. Acad. Sci. USA* 91: 3054-3057). The pharmaceutical preparation of the gene therapy vector can include the gene therapy vector in an acceptable diluent, or can comprise a slow release matrix in which the gene delivery vehicle is imbedded. Alternatively, where the complete gene delivery vector can be produced intact from recombinant cells, *e.g.*, retroviral vectors, the pharmaceutical preparation can include one or more cells that produce the gene delivery system.

The pharmaceutical compositions can be included in a container, pack, or dispenser together with instructions for administration.

Screening and Detection Methods

The isolated nucleic acid molecules of the invention can be used to express MOLX protein (*e.g.*, via a recombinant expression vector in a host cell in gene therapy applications), to detect MOLX mRNA (*e.g.*, in a biological sample) or a genetic lesion in an MOLX gene, and to modulate MOLX activity, as described further, below. In addition, the MOLX proteins can be used to screen drugs or compounds that modulate the MOLX protein activity or expression as well as to treat disorders characterized by insufficient or excessive production of MOLX protein or production of MOLX protein forms that have decreased or aberrant activity compared to MOLX wild-type protein (*e.g.*; diabetes (regulates insulin release); obesity (binds and transport lipids); metabolic disturbances associated with obesity, the metabolic syndrome X as well as anorexia and wasting disorders associated with chronic diseases and various cancers, and infectious disease(possesses anti-microbial activity) and the various dyslipidemias. In addition, the anti-MOLX antibodies of the invention can be used to detect

and isolate MOLX proteins and modulate MOLX activity. In yet a further aspect, the invention can be used in methods to influence appetite, absorption of nutrients and the disposition of metabolic substrates in both a positive and negative fashion.

The invention further pertains to novel agents identified by the screening assays described herein and uses thereof for treatments as described, *supra*.

Screening Assays

The invention provides a method (also referred to herein as a "screening assay") for identifying modulators, *i.e.*, candidate or test compounds or agents (*e.g.*, peptides, peptidomimetics, small molecules or other drugs) that bind to MOLX proteins or have a stimulatory or inhibitory effect on, *e.g.*, MOLX protein expression or MOLX protein activity. The invention also includes compounds identified in the screening assays described herein.

In one embodiment, the invention provides assays for screening candidate or test compounds which bind to or modulate the activity of the membrane-bound form of an MOLX protein or polypeptide or biologically-active portion thereof. The test compounds of the invention can be obtained using any of the numerous approaches in combinatorial library methods known in the art, including: biological libraries; spatially addressable parallel solid phase or solution phase libraries; synthetic library methods requiring deconvolution; the "one-bead one-compound" library method; and synthetic library methods using affinity chromatography selection. The biological library approach is limited to peptide libraries, while the other four approaches are applicable to peptide, non-peptide oligomer or small molecule libraries of compounds. *See, e.g.*, Lam, 1997. *Anticancer Drug Design* 12: 145.

A "small molecule" as used herein, is meant to refer to a composition that has a molecular weight of less than about 5 kD and most preferably less than about 4 kD. Small molecules can be, *e.g.*, nucleic acids, peptides, polypeptides, peptidomimetics, carbohydrates, lipids or other organic or inorganic molecules. Libraries of chemical and/or biological mixtures, such as fungal, bacterial, or algal extracts, are known in the art and can be screened with any of the assays of the invention.

Examples of methods for the synthesis of molecular libraries can be found in the art, for example in: DeWitt, *et al.*, 1993. *Proc. Natl. Acad. Sci. U.S.A.* 90: 6909; Erb, *et al.*, 1994. *Proc. Natl. Acad. Sci. U.S.A.* 91: 11422; Zuckermann, *et al.*, 1994. *J. Med. Chem.* 37: 2678; Cho, *et al.*, 1993. *Science* 261: 1303; Carrell, *et al.*, 1994. *Angew. Chem. Int. Ed. Engl.* 33: 2059; Carell, *et al.*, 1994. *Angew. Chem. Int. Ed. Engl.* 33: 2061; and Gallop, *et al.*, 1994. *J. Med. Chem.* 37: 1233.

Libraries of compounds may be presented in solution (*e.g.*, Houghten, 1992. *Biotechniques* 13: 412-421), or on beads (Lam, 1991. *Nature* 354: 82-84), on chips (Fodor, 1993. *Nature* 364: 555-556), bacteria (Ladner, U.S. Patent No. 5,223,409), spores (Ladner, U.S. Patent 5,233,409), plasmids (Cull, *et al.*, 1992. *Proc. Natl. Acad. Sci. USA* 89: 1865-1869) or on phage (Scott and Smith, 1990. *Science* 249: 386-390; Devlin, 1990. *Science* 249: 404-406; Cwirla, *et al.*, 1990. *Proc. Natl. Acad. Sci. U.S.A.* 87: 6378-6382; Felici, 1991. *J. Mol. Biol.* 222: 301-310; Ladner, U.S. Patent No. 5,233,409.).

In one embodiment, an assay is a cell-based assay in which a cell which expresses a membrane-bound form of MOLX protein, or a biologically-active portion thereof, on the cell surface is contacted with a test compound and the ability of the test compound to bind to an MOLX protein determined. The cell, for example, can of mammalian origin or a yeast cell. Determining the ability of the test compound to bind to the MOLX protein can be accomplished, for example, by coupling the test compound with a radioisotope or enzymatic label such that binding of the test compound to the MOLX protein or biologically-active portion thereof can be determined by detecting the labeled compound in a complex. For example, test compounds can be labeled with ^{125}I, ^{35}S, ^{14}C, or ^{3}H, either directly or indirectly, and the radioisotope detected by direct counting of radioemission or by scintillation counting. Alternatively, test compounds can be enzymatically-labeled with, for example, horseradish peroxidase, alkaline phosphatase, or luciferase, and the enzymatic label detected by determination of conversion of an appropriate substrate to product. In one embodiment, the assay comprises contacting a cell which expresses a membrane-bound form of MOLX protein, or a biologically-active portion thereof, on the cell surface with a known compound which binds MOLX to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with an MOLX protein, wherein determining the ability of the test compound to interact with an MOLX protein comprises determining the ability of the test compound to preferentially bind to MOLX protein or a biologically-active portion thereof as compared to the known compound.

In another embodiment, an assay is a cell-based assay comprising contacting a cell expressing a membrane-bound form of MOLX protein, or a biologically-active portion thereof, on the cell surface with a test compound and determining the ability of the test compound to modulate (*e.g.*, stimulate or inhibit) the activity of the MOLX protein or biologically-active portion thereof. Determining the ability of the test compound to modulate the activity of MOLX or a biologically-active portion thereof can be accomplished, for example, by determining the ability of the MOLX protein to bind to or interact with an MOLX

target molecule. As used herein, a "target molecule" is a molecule with which an MOLX protein binds or interacts in nature, for example, a molecule on the surface of a cell which expresses an MOLX interacting protein, a molecule on the surface of a second cell, a molecule in the extracellular milieu, a molecule associated with the internal surface of a cell membrane or a cytoplasmic molecule. An MOLX target molecule can be a non-MOLX molecule or an MOLX protein or polypeptide of the invention. In one embodiment, an MOLX target molecule is a component of a signal transduction pathway that facilitates transduction of an extracellular signal (*e.g.* a signal generated by binding of a compound to a membrane-bound MOLX molecule) through the cell membrane and into the cell. The target, for example, can be a second intercellular protein that has catalytic activity or a protein that facilitates the association of downstream signaling molecules with MOLX.

Determining the ability of the MOLX protein to bind to or interact with an MOLX target molecule can be accomplished by one of the methods described above for determining direct binding. In one embodiment, determining the ability of the MOLX protein to bind to or interact with an MOLX target molecule can be accomplished by determining the activity of the target molecule. For example, the activity of the target molecule can be determined by detecting induction of a cellular second messenger of the target (*i.e.* intracellular Ca^{2+}, diacylglycerol, IP_3, etc.), detecting catalytic/enzymatic activity of the target an appropriate substrate, detecting the induction of a reporter gene (comprising an MOLX-responsive regulatory element operatively linked to a nucleic acid encoding a detectable marker, *e.g.*, luciferase), or detecting a cellular response, for example, cell survival, cellular differentiation, or cell proliferation.

In yet another embodiment, an assay of the invention is a cell-free assay comprising contacting an MOLX protein or biologically-active portion thereof with a test compound and determining the ability of the test compound to bind to the MOLX protein or biologically-active portion thereof. Binding of the test compound to the MOLX protein can be determined either directly or indirectly as described above. In one such embodiment, the assay comprises contacting the MOLX protein or biologically-active portion thereof with a known compound which binds MOLX to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with an MOLX protein, wherein determining the ability of the test compound to interact with an MOLX protein comprises determining the ability of the test compound to preferentially bind to MOLX or biologically-active portion thereof as compared to the known compound.

In still another embodiment, an assay is a cell-free assay comprising contacting MOLX protein or biologically-active portion thereof with a test compound and determining the ability of the test compound to modulate (*e.g.* stimulate or inhibit) the activity of the MOLX protein or biologically-active portion thereof. Determining the ability of the test compound to modulate the activity of MOLX can be accomplished, for example, by determining the ability of the MOLX protein to bind to an MOLX target molecule by one of the methods described above for determining direct binding. In an alternative embodiment, determining the ability of the test compound to modulate the activity of MOLX protein can be accomplished by determining the ability of the MOLX protein further modulate an MOLX target molecule. For example, the catalytic/enzymatic activity of the target molecule on an appropriate substrate can be determined as described, *supra*.

In yet another embodiment, the cell-free assay comprises contacting the MOLX protein or biologically-active portion thereof with a known compound which binds MOLX protein to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with an MOLX protein, wherein determining the ability of the test compound to interact with an MOLX protein comprises determining the ability of the MOLX protein to preferentially bind to or modulate the activity of an MOLX target molecule.

The cell-free assays of the invention are amenable to use of both the soluble form or the membrane-bound form of MOLX protein. In the case of cell-free assays comprising the membrane-bound form of MOLX protein, it may be desirable to utilize a solubilizing agent such that the membrane-bound form of MOLX protein is maintained in solution. Examples of such solubilizing agents include non-ionic detergents such as n-octylglucoside, n-dodecylglucoside, n-dodecylmaltoside, octanoyl-N-methylglucamide, decanoyl-N-methylglucamide, Triton® X-100, Triton® X-114, Thesit®, Isotridecypoly(ethylene glycol ether)$_n$, N-dodecyl--N,N-dimethyl-3-ammonio-1-propane sulfonate, 3-(3-cholamidopropyl) dimethylamminiol-1-propane sulfonate (CHAPS), or 3-(3-cholamidopropyl)dimethylamminiol-2-hydroxy-1-propane sulfonate (CHAPSO).

In more than one embodiment of the above assay methods of the invention, it may be desirable to immobilize either MOLX protein or its target molecule to facilitate separation of complexed from uncomplexed forms of one or both of the proteins, as well as to accommodate automation of the assay. Binding of a test compound to MOLX protein, or interaction of MOLX protein with a target molecule in the presence and absence of a candidate compound, can be accomplished in any vessel suitable for containing the reactants. Examples of such

vessels include microtiter plates, test tubes, and micro-centrifuge tubes. In one embodiment, a fusion protein can be provided that adds a domain that allows one or both of the proteins to be bound to a matrix. For example, GST-MOLX fusion proteins or GST-target fusion proteins can be adsorbed onto glutathione sepharose beads (Sigma Chemical, St. Louis, MO) or glutathione derivatized microtiter plates, that are then combined with the test compound or the test compound and either the non-adsorbed target protein or MOLX protein, and the mixture is incubated under conditions conducive to complex formation (*e.g.*, at physiological conditions for salt and pH). Following incubation, the beads or microtiter plate wells are washed to remove any unbound components, the matrix immobilized in the case of beads, complex determined either directly or indirectly, for example, as described, *supra*. Alternatively, the complexes can be dissociated from the matrix, and the level of MOLX protein binding or activity determined using standard techniques.

Other techniques for immobilizing proteins on matrices can also be used in the screening assays of the invention. For example, either the MOLX protein or its target molecule can be immobilized utilizing conjugation of biotin and streptavidin. Biotinylated MOLX protein or target molecules can be prepared from biotin-NHS (N-hydroxy-succinimide) using techniques well-known within the art (*e.g.*, biotinylation kit, Pierce Chemicals, Rockford, Ill.), and immobilized in the wells of streptavidin-coated 96 well plates (Pierce Chemical). Alternatively, antibodies reactive with MOLX protein or target molecules, but which do not interfere with binding of the MOLX protein to its target molecule, can be derivatized to the wells of the plate, and unbound target or MOLX protein trapped in the wells by antibody conjugation. Methods for detecting such complexes, in addition to those described above for the GST-immobilized complexes, include immunodetection of complexes using antibodies reactive with the MOLX protein or target molecule, as well as enzyme-linked assays that rely on detecting an enzymatic activity associated with the MOLX protein or target molecule.

In another embodiment, modulators of MOLX protein expression are identified in a method wherein a cell is contacted with a candidate compound and the expression of MOLX mRNA or protein in the cell is determined. The level of expression of MOLX mRNA or protein in the presence of the candidate compound is compared to the level of expression of MOLX mRNA or protein in the absence of the candidate compound. The candidate compound can then be identified as a modulator of MOLX mRNA or protein expression based upon this comparison. For example, when expression of MOLX mRNA or protein is greater (*i.e.*, statistically significantly greater) in the presence of the candidate compound than in its

absence, the candidate compound is identified as a stimulator of MOLX mRNA or protein expression. Alternatively, when expression of MOLX mRNA or protein is less (statistically significantly less) in the presence of the candidate compound than in its absence, the candidate compound is identified as an inhibitor of MOLX mRNA or protein expression. The level of MOLX mRNA or protein expression in the cells can be determined by methods described herein for detecting MOLX mRNA or protein.

In yet another aspect of the invention, the MOLX proteins can be used as "bait proteins" in a two-hybrid assay or three hybrid assay (*see, e.g.*, U.S. Patent No. 5,283,317; Zervos, *et al.*, 1993. *Cell* 72: 223-232; Madura, *et al.*, 1993. *J. Biol. Chem.* 268: 12046-12054; Bartel, *et al.*, 1993. *Biotechniques* 14: 920-924; Iwabuchi, *et al.*, 1993. *Oncogene* 8: 1693-1696; and Brent WO 94/10300), to identify other proteins that bind to or interact with MOLX ("MOLX-binding proteins" or "MOLX-bp") and modulate MOLX activity. Such MOLX-binding proteins are also likely to be involved in the propagation of signals by the MOLX proteins as, for example, upstream or downstream elements of the MOLX pathway.

The two-hybrid system is based on the modular nature of most transcription factors, which consist of separable DNA-binding and activation domains. Briefly, the assay utilizes two different DNA constructs. In one construct, the gene that codes for MOLX is fused to a gene encoding the DNA binding domain of a known transcription factor (*e.g.*, GAL-4). In the other construct, a DNA sequence, from a library of DNA sequences, that encodes an unidentified protein ("prey" or "sample") is fused to a gene that codes for the activation domain of the known transcription factor. If the "bait" and the "prey" proteins are able to interact, *in vivo*, forming an MOLX-dependent complex, the DNA-binding and activation domains of the transcription factor are brought into close proximity. This proximity allows transcription of a reporter gene (*e.g.*, LacZ) that is operably linked to a transcriptional regulatory site responsive to the transcription factor. Expression of the reporter gene can be detected and cell colonies containing the functional transcription factor can be isolated and used to obtain the cloned gene that encodes the protein which interacts with MOLX.

The invention further pertains to novel agents identified by the aforementioned screening assays and uses thereof for treatments as described herein.

Detection Assays

Portions or fragments of the cDNA sequences identified herein (and the corresponding complete gene sequences) can be used in numerous ways as polynucleotide reagents. By way of example, and not of limitation, these sequences can be used to: (*i*) map their respective

genes on a chromosome; and, thus, locate gene regions associated with genetic disease; (*ii*) identify an individual from a minute biological sample (tissue typing); and (*iii*) aid in forensic identification of a biological sample. Some of these applications are described in the subsections, below.

Chromosome Mapping

Once the sequence (or a portion of the sequence) of a gene has been isolated, this sequence can be used to map the location of the gene on a chromosome. This process is called chromosome mapping. Accordingly, portions or fragments of the MOLX sequences, SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, and 23, or fragments or derivatives thereof, can be used to map the location of the MOLX genes, respectively, on a chromosome. The mapping of the MOLX sequences to chromosomes is an important first step in correlating these sequences with genes associated with disease.

Briefly, MOLX genes can be mapped to chromosomes by preparing PCR primers (preferably 15-25 bp in length) from the MOLX sequences. Computer analysis of the MOLX, sequences can be used to rapidly select primers that do not span more than one exon in the genomic DNA, thus complicating the amplification process. These primers can then be used for PCR screening of somatic cell hybrids containing individual human chromosomes. Only those hybrids containing the human gene corresponding to the MOLX sequences will yield an amplified fragment.

Somatic cell hybrids are prepared by fusing somatic cells from different mammals (*e.g.*, human and mouse cells). As hybrids of human and mouse cells grow and divide, they gradually lose human chromosomes in random order, but retain the mouse chromosomes. By using media in which mouse cells cannot grow, because they lack a particular enzyme, but in which human cells can, the one human chromosome that contains the gene encoding the needed enzyme will be retained. By using various media, panels of hybrid cell lines can be established. Each cell line in a panel contains either a single human chromosome or a small number of human chromosomes, and a full set of mouse chromosomes, allowing easy mapping of individual genes to specific human chromosomes. *See, e.g.,* D'Eustachio, *et al.*, 1983. *Science* 220: 919-924. Somatic cell hybrids containing only fragments of human chromosomes can also be produced by using human chromosomes with translocations and deletions.

PCR mapping of somatic cell hybrids is a rapid procedure for assigning a particular sequence to a particular chromosome. Three or more sequences can be assigned per day using

a single thermal cycler. Using the MOLX sequences to design oligonucleotide primers, sub-localization can be achieved with panels of fragments from specific chromosomes.

Fluorescence *in situ* hybridization (FISH) of a DNA sequence to a metaphase chromosomal spread can further be used to provide a precise chromosomal location in one step. Chromosome spreads can be made using cells whose division has been blocked in metaphase by a chemical like colcemid that disrupts the mitotic spindle. The chromosomes can be treated briefly with trypsin, and then stained with Giemsa. A pattern of light and dark bands develops on each chromosome, so that the chromosomes can be identified individually. The FISH technique can be used with a DNA sequence as short as 500 or 600 bases. However, clones larger than 1,000 bases have a higher likelihood of binding to a unique chromosomal location with sufficient signal intensity for simple detection. Preferably 1,000 bases, and more preferably 2,000 bases, will suffice to get good results at a reasonable amount of time. For a review of this technique, *see,* Verma, *et al.*, HUMAN CHROMOSOMES: A MANUAL OF BASIC TECHNIQUES (Pergamon Press, New York 1988).

Reagents for chromosome mapping can be used individually to mark a single chromosome or a single site on that chromosome, or panels of reagents can be used for marking multiple sites and/or multiple chromosomes. Reagents corresponding to noncoding regions of the genes actually are preferred for mapping purposes. Coding sequences are more likely to be conserved within gene families, thus increasing the chance of cross hybridizations during chromosomal mapping.

Once a sequence has been mapped to a precise chromosomal location, the physical position of the sequence on the chromosome can be correlated with genetic map data. Such data are found, *e.g.*, in McKusick, MENDELIAN INHERITANCE IN MAN, available on-line through Johns Hopkins University Welch Medical Library). The relationship between genes and disease, mapped to the same chromosomal region, can then be identified through linkage analysis (co-inheritance of physically adjacent genes), described in, *e.g.*, Egeland, *et al.*, 1987. *Nature*, 325: 783-787.

Moreover, differences in the DNA sequences between individuals affected and unaffected with a disease associated with the MOLX gene, can be determined. If a mutation is observed in some or all of the affected individuals but not in any unaffected individuals, then the mutation is likely to be the causative agent of the particular disease. Comparison of affected and unaffected individuals generally involves first looking for structural alterations in the chromosomes, such as deletions or translocations that are visible from chromosome spreads or detectable using PCR based on that DNA sequence. Ultimately, complete

sequencing of genes from several individuals can be performed to confirm the presence of a mutation and to distinguish mutations from polymorphisms.

Tissue Typing

The MOLX sequences of the invention can also be used to identify individuals from minute biological samples. In this technique, an individual's genomic DNA is digested with one or more restriction enzymes, and probed on a Southern blot to yield unique bands for identification. The sequences of the invention are useful as additional DNA markers for RFLP ("restriction fragment length polymorphisms," described in U.S. Patent No. 5,272,057).

Furthermore, the sequences of the invention can be used to provide an alternative technique that determines the actual base-by-base DNA sequence of selected portions of an individual's genome. Thus, the MOLX sequences described herein can be used to prepare two PCR primers from the 5'- and 3'-termini of the sequences. These primers can then be used to amplify an individual's DNA and subsequently sequence it.

Panels of corresponding DNA sequences from individuals, prepared in this manner, can provide unique individual identifications, as each individual will have a unique set of such DNA sequences due to allelic differences. The sequences of the invention can be used to obtain such identification sequences from individuals and from tissue. The MOLX sequences of the invention uniquely represent portions of the human genome. Allelic variation occurs to some degree in the coding regions of these sequences, and to a greater degree in the noncoding regions. It is estimated that allelic variation between individual humans occurs with a frequency of about once per each 500 bases. Much of the allelic variation is due to single nucleotide polymorphisms (SNPs), which include restriction fragment length polymorphisms (RFLPs).

Each of the sequences described herein can, to some degree, be used as a standard against which DNA from an individual can be compared for identification purposes. Because greater numbers of polymorphisms occur in the noncoding regions, fewer sequences are necessary to differentiate individuals. The noncoding sequences can comfortably provide positive individual identification with a panel of perhaps 10 to 1,000 primers that each yield a noncoding amplified sequence of 100 bases. If predicted coding sequences, such as those in SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, and 23 are used, a more appropriate number of primers for positive individual identification would be 500-2,000.

Predictive Medicine

The invention also pertains to the field of predictive medicine in which diagnostic assays, prognostic assays, pharmacogenomics, and monitoring clinical trials are used for prognostic (predictive) purposes to thereby treat an individual prophylactically. Accordingly, one aspect of the invention relates to diagnostic assays for determining MOLX protein and/or nucleic acid expression as well as MOLX activity, in the context of a biological sample (*e.g.*, blood, serum, cells, tissue) to thereby determine whether an individual is afflicted with a disease or disorder, or is at risk of developing a disorder, associated with aberrant MOLX expression or activity. The disorders include metabolic disorders, diabetes, obesity, infectious disease, anorexia, cancer-associated cachexia, cancer, neurodegenerative disorders, Alzheimer's Disease, Parkinson's Disorder, immune disorders, and hematopoietic disorders, and the various dyslipidemias, metabolic disturbances associated with obesity, the metabolic syndrome X and wasting disorders associated with chronic diseases and various cancers. The invention also provides for prognostic (or predictive) assays for determining whether an individual is at risk of developing a disorder associated with MOLX protein, nucleic acid expression or activity. For example, mutations in an MOLX gene can be assayed in a biological sample. Such assays can be used for prognostic or predictive purpose to thereby prophylactically treat an individual prior to the onset of a disorder characterized by or associated with MOLX protein, nucleic acid expression, or biological activity.

Another aspect of the invention provides methods for determining MOLX protein, nucleic acid expression or activity in an individual to thereby select appropriate therapeutic or prophylactic agents for that individual (referred to herein as "pharmacogenomics"). Pharmacogenomics allows for the selection of agents (*e.g.*, drugs) for therapeutic or prophylactic treatment of an individual based on the genotype of the individual (*e.g.*, the genotype of the individual examined to determine the ability of the individual to respond to a particular agent.)

Yet another aspect of the invention pertains to monitoring the influence of agents (*e.g.*, drugs, compounds) on the expression or activity of MOLX in clinical trials.

These and other agents are described in further detail in the following sections.

Diagnostic Assays

An exemplary method for detecting the presence or absence of MOLX in a biological sample involves obtaining a biological sample from a test subject and contacting the biological

sample with a compound or an agent capable of detecting MOLX protein or nucleic acid (*e.g.*, mRNA, genomic DNA) that encodes MOLX protein such that the presence of MOLX is detected in the biological sample. An agent for detecting MOLX mRNA or genomic DNA is a labeled nucleic acid probe capable of hybridizing to MOLX mRNA or genomic DNA. The nucleic acid probe can be, for example, a full-length MOLX nucleic acid, such as the nucleic acid of SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, and 23, or a portion thereof, such as an oligonucleotide of at least 15, 30, 50, 100, 250 or 500 nucleotides in length and sufficient to specifically hybridize under stringent conditions to MOLX mRNA or genomic DNA. Other suitable probes for use in the diagnostic assays of the invention are described herein.

An agent for detecting MOLX protein is an antibody capable of binding to MOLX protein, preferably an antibody with a detectable label. Antibodies can be polyclonal, or more preferably, monoclonal. An intact antibody, or a fragment thereof (*e.g.*, Fab or $F(ab')_2$) can be used. The term "labeled", with regard to the probe or antibody, is intended to encompass direct labeling of the probe or antibody by coupling (*i.e.*, physically linking) a detectable substance to the probe or antibody, as well as indirect labeling of the probe or antibody by reactivity with another reagent that is directly labeled. Examples of indirect labeling include detection of a primary antibody using a fluorescently-labeled secondary antibody and end-labeling of a DNA probe with biotin such that it can be detected with fluorescently-labeled streptavidin. The term "biological sample" is intended to include tissues, cells and biological fluids isolated from a subject, as well as tissues, cells and fluids present within a subject. That is, the detection method of the invention can be used to detect MOLX mRNA, protein, or genomic DNA in a biological sample *in vitro* as well as *in vivo*. For example, *in vitro* techniques for detection of MOLX mRNA include Northern hybridizations and *in situ* hybridizations. *In vitro* techniques for detection of MOLX protein include enzyme linked immunosorbent assays (ELISAs), Western blots, immunoprecipitations, and immunofluorescence. *In vitro* techniques for detection of MOLX genomic DNA include Southern hybridizations. Furthermore, *in vivo* techniques for detection of MOLX protein include introducing into a subject a labeled anti-MOLX antibody. For example, the antibody can be labeled with a radioactive marker whose presence and location in a subject can be detected by standard imaging techniques.

In one embodiment, the biological sample contains protein molecules from the test subject. Alternatively, the biological sample can contain mRNA molecules from the test subject or genomic DNA molecules from the test subject. A preferred biological sample is a peripheral blood leukocyte sample isolated by conventional means from a subject.

In another embodiment, the methods further involve obtaining a control biological sample from a control subject, contacting the control sample with a compound or agent capable of detecting MOLX protein, mRNA, or genomic DNA, such that the presence of MOLX protein, mRNA or genomic DNA is detected in the biological sample, and comparing the presence of MOLX protein, mRNA or genomic DNA in the control sample with the presence of MOLX protein, mRNA or genomic DNA in the test sample.

The invention also encompasses kits for detecting the presence of MOLX in a biological sample. For example, the kit can comprise: a labeled compound or agent capable of detecting MOLX protein or mRNA in a biological sample; means for determining the amount of MOLX in the sample; and means for comparing the amount of MOLX in the sample with a standard. The compound or agent can be packaged in a suitable container. The kit can further comprise instructions for using the kit to detect MOLX protein or nucleic acid.

Prognostic Assays

The diagnostic methods described herein can furthermore be utilized to identify subjects having or at risk of developing a disease or disorder associated with aberrant MOLX expression or activity. For example, the assays described herein, such as the preceding diagnostic assays or the following assays, can be utilized to identify a subject having or at risk of developing a disorder associated with MOLX protein, nucleic acid expression or activity. Alternatively, the prognostic assays can be utilized to identify a subject having or at risk for developing a disease or disorder. Thus, the invention provides a method for identifying a disease or disorder associated with aberrant MOLX expression or activity in which a test sample is obtained from a subject and MOLX protein or nucleic acid (*e.g.*, mRNA, genomic DNA) is detected, wherein the presence of MOLX protein or nucleic acid is diagnostic for a subject having or at risk of developing a disease or disorder associated with aberrant MOLX expression or activity. As used herein, a "test sample" refers to a biological sample obtained from a subject of interest. For example, a test sample can be a biological fluid (*e.g.*, serum), cell sample, or tissue.

Furthermore, the prognostic assays described herein can be used to determine whether a subject can be administered an agent (*e.g.*, an agonist, antagonist, peptidomimetic, protein, peptide, nucleic acid, small molecule, or other drug candidate) to treat a disease or disorder associated with aberrant MOLX expression or activity. For example, such methods can be used to determine whether a subject can be effectively treated with an agent for a disorder. Thus, the invention provides methods for determining whether a subject can be effectively

treated with an agent for a disorder associated with aberrant MOLX expression or activity in which a test sample is obtained and MOLX protein or nucleic acid is detected (*e.g.*, wherein the presence of MOLX protein or nucleic acid is diagnostic for a subject that can be administered the agent to treat a disorder associated with aberrant MOLX expression or activity).

The methods of the invention can also be used to detect genetic lesions in an MOLX gene, thereby determining if a subject with the lesioned gene is at risk for a disorder characterized by aberrant cell proliferation and/or differentiation. In various embodiments, the methods include detecting, in a sample of cells from the subject, the presence or absence of a genetic lesion characterized by at least one of an alteration affecting the integrity of a gene encoding an MOLX-protein, or the misexpression of the MOLX gene. For example, such genetic lesions can be detected by ascertaining the existence of at least one of: (*i*) a deletion of one or more nucleotides from an MOLX gene; (*ii*) an addition of one or more nucleotides to an MOLX gene; (*iii*) a substitution of one or more nucleotides of an MOLX gene, (*iv*) a chromosomal rearrangement of an MOLX gene; (*v*) an alteration in the level of a messenger RNA transcript of an MOLX gene, (*vi*) aberrant modification of an MOLX gene, such as of the methylation pattern of the genomic DNA, (*vii*) the presence of a non-wild-type splicing pattern of a messenger RNA transcript of an MOLX gene, (*viii*) a non-wild-type level of an MOLX protein, (*ix*) allelic loss of an MOLX gene, and (*x*) inappropriate post-translational modification of an MOLX protein. As described herein, there are a large number of assay techniques known in the art which can be used for detecting lesions in an MOLX gene. A preferred biological sample is a peripheral blood leukocyte sample isolated by conventional means from a subject. However, any biological sample containing nucleated cells may be used, including, for example, buccal mucosal cells.

In certain embodiments, detection of the lesion involves the use of a probe/primer in a polymerase chain reaction (PCR) (*see, e.g.*, U.S. Patent Nos. 4,683,195 and 4,683,202), such as anchor PCR or RACE PCR, or, alternatively, in a ligation chain reaction (LCR) (*see, e.g.*, Landegran, *et al.*, 1988. *Science* 241: 1077-1080; and Nakazawa, *et al.*, 1994. *Proc. Natl. Acad. Sci. USA* 91: 360-364), the latter of which can be particularly useful for detecting point mutations in the MOLX-gene (*see,* Abravaya, *et al.*, 1995. *Nucl. Acids Res.* 23: 675-682). This method can include the steps of collecting a sample of cells from a patient, isolating nucleic acid (*e.g.*, genomic, mRNA or both) from the cells of the sample, contacting the nucleic acid sample with one or more primers that specifically hybridize to an MOLX gene under conditions such that hybridization and amplification of the MOLX gene (if present)

occurs, and detecting the presence or absence of an amplification product, or detecting the size of the amplification product and comparing the length to a control sample. It is anticipated that PCR and/or LCR may be desirable to use as a preliminary amplification step in conjunction with any of the techniques used for detecting mutations described herein.

Alternative amplification methods include: self sustained sequence replication (*see,* Guatelli, *et al.*, 1990. *Proc. Natl. Acad. Sci. USA* 87: 1874-1878), transcriptional amplification system (*see,* Kwoh, *et al.*, 1989. *Proc. Natl. Acad. Sci. USA* 86: 1173-1177); Qβ Replicase (*see,* Lizardi, *et al*, 1988. *BioTechnology* 6: 1197), or any other nucleic acid amplification method, followed by the detection of the amplified molecules using techniques well known to those of skill in the art. These detection schemes are especially useful for the detection of nucleic acid molecules if such molecules are present in very low numbers.

In an alternative embodiment, mutations in an MOLX gene from a sample cell can be identified by alterations in restriction enzyme cleavage patterns. For example, sample and control DNA is isolated, amplified (optionally), digested with one or more restriction endonucleases, and fragment length sizes are determined by gel electrophoresis and compared. Differences in fragment length sizes between sample and control DNA indicates mutations in the sample DNA. Moreover, the use of sequence specific ribozymes (*see, e.g.,* U.S. Patent No. 5,493,531) can be used to score for the presence of specific mutations by development or loss of a ribozyme cleavage site.

In other embodiments, genetic mutations in MOLX can be identified by hybridizing a sample and control nucleic acids, *e.g.*, DNA or RNA, to high-density arrays containing hundreds or thousands of oligonucleotides probes. *See, e.g.,* Cronin, *et al.*, 1996. *Human Mutation* 7: 244-255; Kozal, *et al.*, 1996. *Nat. Med.* 2: 753-759. For example, genetic mutations in MOLX can be identified in two dimensional arrays containing light-generated DNA probes as described in Cronin, *et al.*, *supra.* Briefly, a first hybridization array of probes can be used to scan through long stretches of DNA in a sample and control to identify base changes between the sequences by making linear arrays of sequential overlapping probes. This step allows the identification of point mutations. This is followed by a second hybridization array that allows the characterization of specific mutations by using smaller, specialized probe arrays complementary to all variants or mutations detected. Each mutation array is composed of parallel probe sets, one complementary to the wild-type gene and the other complementary to the mutant gene.

In yet another embodiment, any of a variety of sequencing reactions known in the art can be used to directly sequence the MOLX gene and detect mutations by comparing the

sequence of the sample MOLX with the corresponding wild-type (control) sequence. Examples of sequencing reactions include those based on techniques developed by Maxim and Gilbert, 1977. *Proc. Natl. Acad. Sci. USA* 74: 560 or Sanger, 1977. *Proc. Natl. Acad. Sci. USA* 74: 5463. It is also contemplated that any of a variety of automated sequencing procedures can be utilized when performing the diagnostic assays (*see, e.g.,* Naeve, *et al.*, 1995. *Biotechniques* 19: 448), including sequencing by mass spectrometry (see, *e.g.*, PCT International Publication No. WO 94/16101; Cohen, *et al.*, 1996. *Adv. Chromatography* 36: 127-162; and Griffin, *et al.*, 1993. *Appl. Biochem. Biotechnol.* 38: 147-159).

Other methods for detecting mutations in the MOLX gene include methods in which protection from cleavage agents is used to detect mismatched bases in RNA/RNA or RNA/DNA heteroduplexes. *See, e.g.,* Myers, *et al.*, 1985. *Science* 230: 1242. In general, the art technique of "mismatch cleavage" starts by providing heteroduplexes of formed by hybridizing (labeled) RNA or DNA containing the wild-type MOLX sequence with potentially mutant RNA or DNA obtained from a tissue sample. The double-stranded duplexes are treated with an agent that cleaves single-stranded regions of the duplex such as which will exist due to basepair mismatches between the control and sample strands. For instance, RNA/DNA duplexes can be treated with RNase and DNA/DNA hybrids treated with S_1 nuclease to enzymatically digesting the mismatched regions. In other embodiments, either DNA/DNA or RNA/DNA duplexes can be treated with hydroxylamine or osmium tetroxide and with piperidine in order to digest mismatched regions. After digestion of the mismatched regions, the resulting material is then separated by size on denaturing polyacrylamide gels to determine the site of mutation. *See, e.g.,* Cotton, *et al.*, 1988. *Proc. Natl. Acad. Sci. USA* 85: 4397; Saleeba, *et al.*, 1992. *Methods Enzymol.* 217: 286-295. In an embodiment, the control DNA or RNA can be labeled for detection.

In still another embodiment, the mismatch cleavage reaction employs one or more proteins that recognize mismatched base pairs in double-stranded DNA (so called "DNA mismatch repair" enzymes) in defined systems for detecting and mapping point mutations in MOLX cDNAs obtained from samples of cells. For example, the mutY enzyme of *E. coli* cleaves A at G/A mismatches and the thymidine DNA glycosylase from HeLa cells cleaves T at G/T mismatches. *See, e.g.,* Hsu, *et al.*, 1994. *Carcinogenesis* 15: 1657-1662. According to an exemplary embodiment, a probe based on an MOLX sequence, *e.g.*, a wild-type MOLX sequence, is hybridized to a cDNA or other DNA product from a test cell(s). The duplex is treated with a DNA mismatch repair enzyme, and the cleavage products, if any, can be detected from electrophoresis protocols or the like. *See, e.g.,* U.S. Patent No. 5,459,039.

In other embodiments, alterations in electrophoretic mobility will be used to identify mutations in MOLX genes. For example, single strand conformation polymorphism (SSCP) may be used to detect differences in electrophoretic mobility between mutant and wild type nucleic acids. *See, e.g.,* Orita, *et al.*, 1989. *Proc. Natl. Acad. Sci. USA*: 86: 2766; Cotton, 1993. *Mutat. Res.* 285: 125-144; Hayashi, 1992. *Genet. Anal. Tech. Appl.* 9: 73-79. Single-stranded DNA fragments of sample and control MOLX nucleic acids will be denatured and allowed to renature. The secondary structure of single-stranded nucleic acids varies according to sequence, the resulting alteration in electrophoretic mobility enables the detection of even a single base change. The DNA fragments may be labeled or detected with labeled probes. The sensitivity of the assay may be enhanced by using RNA (rather than DNA), in which the secondary structure is more sensitive to a change in sequence. In one embodiment, the subject method utilizes heteroduplex analysis to separate double stranded heteroduplex molecules on the basis of changes in electrophoretic mobility. *See, e.g.,* Keen, *et al.*, 1991. *Trends Genet.* 7: 5.

In yet another embodiment, the movement of mutant or wild-type fragments in polyacrylamide gels containing a gradient of denaturant is assayed using denaturing gradient gel electrophoresis (DGGE). *See, e.g.,* Myers, *et al.*, 1985. *Nature* 313: 495. When DGGE is used as the method of analysis, DNA will be modified to insure that it does not completely denature, for example by adding a GC clamp of approximately 40 bp of high-melting GC-rich DNA by PCR. In a further embodiment, a temperature gradient is used in place of a denaturing gradient to identify differences in the mobility of control and sample DNA. *See, e.g.,* Rosenbaum and Reissner, 1987. *Biophys. Chem.* 265: 12753.

Examples of other techniques for detecting point mutations include, but are not limited to, selective oligonucleotide hybridization, selective amplification, or selective primer extension. For example, oligonucleotide primers may be prepared in which the known mutation is placed centrally and then hybridized to target DNA under conditions that permit hybridization only if a perfect match is found. *See, e.g.,* Saiki, *et al.*, 1986. *Nature* 324: 163; Saiki, *et al.*, 1989. *Proc. Natl. Acad. Sci. USA* 86: 6230. Such allele specific oligonucleotides are hybridized to PCR amplified target DNA or a number of different mutations when the oligonucleotides are attached to the hybridizing membrane and hybridized with labeled target DNA.

Alternatively, allele specific amplification technology that depends on selective PCR amplification may be used in conjunction with the instant invention. Oligonucleotides used as primers for specific amplification may carry the mutation of interest in the center of the

molecule (so that amplification depends on differential hybridization; *see, e.g.,* Gibbs, *et al.*, 1989. *Nucl. Acids Res.* 17: 2437-2448) or at the extreme 3'-terminus of one primer where, under appropriate conditions, mismatch can prevent, or reduce polymerase extension (*see, e.g.,* Prossner, 1993. *Tibtech.* 11: 238). In addition it may be desirable to introduce a novel restriction site in the region of the mutation to create cleavage-based detection. *See, e.g.,* Gasparini, *et al.,* 1992. *Mol. Cell Probes* 6: 1. It is anticipated that in certain embodiments amplification may also be performed using *Taq* ligase for amplification. *See, e.g.,* Barany, 1991. *Proc. Natl. Acad. Sci. USA* 88: 189. In such cases, ligation will occur only if there is a perfect match at the 3'-terminus of the 5' sequence, making it possible to detect the presence of a known mutation at a specific site by looking for the presence or absence of amplification.

The methods described herein may be performed, for example, by utilizing pre-packaged diagnostic kits comprising at least one probe nucleic acid or antibody reagent described herein, which may be conveniently used, *e.g.*, in clinical settings to diagnose patients exhibiting symptoms or family history of a disease or illness involving an MOLX gene.

Furthermore, any cell type or tissue, preferably peripheral blood leukocytes, in which MOLX is expressed may be utilized in the prognostic assays described herein. However, any biological sample containing nucleated cells may be used, including, for example, buccal mucosal cells.

Pharmacogenomics

Agents, or modulators that have a stimulatory or inhibitory effect on MOLX activity (*e.g.*, MOLX gene expression), as identified by a screening assay described herein can be administered to individuals to treat (prophylactically or therapeutically) disorders (The disorders include metabolic disorders, diabetes, obesity, infectious disease, anorexia, cancer-associated cachexia, cancer, neurodegenerative disorders, Alzheimer's Disease, Parkinson's Disorder, immune disorders, and hematopoietic disorders, and the various dyslipidemias, metabolic disturbances associated with obesity, the metabolic syndrome X and wasting disorders associated with chronic diseases and various cancers.) In conjunction with such treatment, the pharmacogenomics (*i.e.*, the study of the relationship between an individual's genotype and that individual's response to a foreign compound or drug) of the individual may be considered. Differences in metabolism of therapeutics can lead to severe toxicity or therapeutic failure by altering the relation between dose and blood concentration of the pharmacologically active drug. Thus, the pharmacogenomics of the individual permits the

selection of effective agents (*e.g.*, drugs) for prophylactic or therapeutic treatments based on a consideration of the individual's genotype. Such pharmacogenomics can further be used to determine appropriate dosages and therapeutic regimens. Accordingly, the activity of MOLX protein, expression of MOLX nucleic acid, or mutation content of MOLX genes in an individual can be determined to thereby select appropriate agent(s) for therapeutic or prophylactic treatment of the individual.

Pharmacogenomics deals with clinically significant hereditary variations in the response to drugs due to altered drug disposition and abnormal action in affected persons. See *e.g.*, Eichelbaum, 1996. *Clin. Exp. Pharmacol. Physiol.*, 23: 983-985; Linder, 1997. *Clin. Chem.*, 43: 254-266. In general, two types of pharmacogenetic conditions can be differentiated. Genetic conditions transmitted as a single factor altering the way drugs act on the body (altered drug action) or genetic conditions transmitted as single factors altering the way the body acts on drugs (altered drug metabolism). These pharmacogenetic conditions can occur either as rare defects or as polymorphisms. For example, glucose-6-phosphate dehydrogenase (G6PD) deficiency is a common inherited enzymopathy in which the main clinical complication is hemolysis after ingestion of oxidant drugs (anti-malarials, sulfonamides, analgesics, nitrofurans) and consumption of fava beans.

As an illustrative embodiment, the activity of drug metabolizing enzymes is a major determinant of both the intensity and duration of drug action. The discovery of genetic polymorphisms of drug metabolizing enzymes (*e.g.*, N-acetyltransferase 2 (NAT 2) and cytochrome P450 enzymes CYP2D6 and CYP2C19) has provided an explanation as to why some patients do not obtain the expected drug effects or show exaggerated drug response and serious toxicity after taking the standard and safe dose of a drug. These polymorphisms are expressed in two phenotypes in the population, the extensive metabolizer (EM) and poor metabolizer (PM). The prevalence of PM is different among different populations. For example, the gene coding for CYP2D6 is highly polymorphic and several mutations have been identified in PM, which all lead to the absence of functional CYP2D6. Poor metabolizers of CYP2D6 and CYP2C19 quite frequently experience exaggerated drug response and side effects when they receive standard doses. If a metabolite is the active therapeutic moiety, PM show no therapeutic response, as demonstrated for the analgesic effect of codeine mediated by its CYP2D6-formed metabolite morphine. At the other extreme are the so called ultra-rapid metabolizers who do not respond to standard doses. Recently, the molecular basis of ultra-rapid metabolism has been identified to be due to CYP2D6 gene amplification.

Thus, the activity of MOLX protein, expression of MOLX nucleic acid, or mutation content of MOLX genes in an individual can be determined to thereby select appropriate agent(s) for therapeutic or prophylactic treatment of the individual. In addition, pharmacogenetic studies can be used to apply genotyping of polymorphic alleles encoding drug-metabolizing enzymes to the identification of an individual's drug responsiveness phenotype. This knowledge, when applied to dosing or drug selection, can avoid adverse reactions or therapeutic failure and thus enhance therapeutic or prophylactic efficiency when treating a subject with an MOLX modulator, such as a modulator identified by one of the exemplary screening assays described herein.

Monitoring of Effects During Clinical Trials

Monitoring the influence of agents (*e.g.*, drugs, compounds) on the expression or activity of MOLX (*e.g.*, the ability to modulate aberrant cell proliferation and/or differentiation) can be applied not only in basic drug screening, but also in clinical trials. For example, the effectiveness of an agent determined by a screening assay as described herein to increase MOLX gene expression, protein levels, or upregulate MOLX activity, can be monitored in clinical trails of subjects exhibiting decreased MOLX gene expression, protein levels, or downregulated MOLX activity. Alternatively, the effectiveness of an agent determined by a screening assay to decrease MOLX gene expression, protein levels, or downregulate MOLX activity, can be monitored in clinical trails of subjects exhibiting increased MOLX gene expression, protein levels, or upregulated MOLX activity. In such clinical trials, the expression or activity of MOLX and, preferably, other genes that have been implicated in, for example, a cellular proliferation or immune disorder can be used as a "read out" or markers of the immune responsiveness of a particular cell.

By way of example, and not of limitation, genes, including MOLX, that are modulated in cells by treatment with an agent (*e.g.*, compound, drug or small molecule) that modulates MOLX activity (*e.g.*, identified in a screening assay as described herein) can be identified. Thus, to study the effect of agents on cellular proliferation disorders, for example, in a clinical trial, cells can be isolated and RNA prepared and analyzed for the levels of expression of MOLX and other genes implicated in the disorder. The levels of gene expression (*i.e.*, a gene expression pattern) can be quantified by Northern blot analysis or RT-PCR, as described herein, or alternatively by measuring the amount of protein produced, by one of the methods as described herein, or by measuring the levels of activity of MOLX or other genes. In this manner, the gene expression pattern can serve as a marker, indicative of the physiological

response of the cells to the agent. Accordingly, this response state may be determined before, and at various points during, treatment of the individual with the agent.

In one embodiment, the invention provides a method for monitoring the effectiveness of treatment of a subject with an agent (*e.g.*, an agonist, antagonist, protein, peptide, peptidomimetic, nucleic acid, small molecule, or other drug candidate identified by the screening assays described herein) comprising the steps of (*i*) obtaining a pre-administration sample from a subject prior to administration of the agent; (*ii*) detecting the level of expression of an MOLX protein, mRNA, or genomic DNA in the preadministration sample; (*iii*) obtaining one or more post-administration samples from the subject; (*iv*) detecting the level of expression or activity of the MOLX protein, mRNA, or genomic DNA in the post-administration samples; (*v*) comparing the level of expression or activity of the MOLX protein, mRNA, or genomic DNA in the pre-administration sample with the MOLX protein, mRNA, or genomic DNA in the post administration sample or samples; and (*vi*) altering the administration of the agent to the subject accordingly. For example, increased administration of the agent may be desirable to increase the expression or activity of MOLX to higher levels than detected, *i.e.*, to increase the effectiveness of the agent. Alternatively, decreased administration of the agent may be desirable to decrease expression or activity of MOLX to lower levels than detected, *i.e.*, to decrease the effectiveness of the agent.

Methods of Treatment

The invention provides for both prophylactic and therapeutic methods of treating a subject at risk of (or susceptible to) a disorder or having a disorder associated with aberrant MOLX expression or activity. The disorders include cardiomyopathy, atherosclerosis, hypertension, congenital heart defects, aortic stenosis, atrial septal defect (ASD), atrioventricular (A-V) canal defect, ductus arteriosus, pulmonary stenosis, subaortic stenosis, ventricular septal defect (VSD), valve diseases, tuberous sclerosis, scleroderma, obesity, transplantation, adrenoleukodystrophy, congenital adrenal hyperplasia, prostate cancer, neoplasm; adenocarcinoma, lymphoma, uterus cancer, fertility, hemophilia, hypercoagulation, idiopathic thrombocytopenic purpura, immunodeficiencies, graft versus host disease, AIDS, bronchial asthma, Crohn's disease; multiple sclerosis, treatment of Albright Hereditary Ostoeodystrophy, and other diseases, disorders and conditions of the like.

These methods of treatment will be discussed more fully, below.

Disease and Disorders

Diseases and disorders that are characterized by increased (relative to a subject not suffering from the disease or disorder) levels or biological activity may be treated with Therapeutics that antagonize (*i.e.*, reduce or inhibit) activity. Therapeutics that antagonize activity may be administered in a therapeutic or prophylactic manner. Therapeutics that may be utilized include, but are not limited to: (*i*) an aforementioned peptide, or analogs, derivatives, fragments or homologs thereof; (*ii*) antibodies to an aforementioned peptide; (*iii*) nucleic acids encoding an aforementioned peptide; (*iv*) administration of antisense nucleic acid and nucleic acids that are "dysfunctional" (*i.e.*, due to a heterologous insertion within the coding sequences of coding sequences to an aforementioned peptide) that are utilized to "knockout" endoggenous function of an aforementioned peptide by homologous recombination (*see, e.g.*, Capecchi, 1989. *Science* 244: 1288-1292); or (*v*) modulators (*i.e.*, inhibitors, agonists and antagonists, including additional peptide mimetic of the invention or antibodies specific to a peptide of the invention) that alter the interaction between an aforementioned peptide and its binding partner.

Diseases and disorders that are characterized by decreased (relative to a subject not suffering from the disease or disorder) levels or biological activity may be treated with Therapeutics that increase (*i.e.*, are agonists to) activity. Therapeutics that upregulate activity may be administered in a therapeutic or prophylactic manner. Therapeutics that may be utilized include, but are not limited to, an aforementioned peptide, or analogs, derivatives, fragments or homologs thereof; or an agonist that increases bioavailability.

Increased or decreased levels can be readily detected by quantifying peptide and/or RNA, by obtaining a patient tissue sample (*e.g.*, from biopsy tissue) and assaying it *in vitro* for RNA or peptide levels, structure and/or activity of the expressed peptides (or mRNAs of an aforementioned peptide). Methods that are well-known within the art include, but are not limited to, immunoassays (*e.g.*, by Western blot analysis, immunoprecipitation followed by sodium dodecyl sulfate (SDS) polyacrylamide gel electrophoresis, immunocytochemistry, etc.) and/or hybridization assays to detect expression of mRNAs (*e.g.*, Northern assays, dot blots, *in situ* hybridization, and the like).

Prophylactic Methods

In one aspect, the invention provides a method for preventing, in a subject, a disease or condition associated with an aberrant MOLX expression or activity, by administering to the

subject an agent that modulates MOLX expression or at least one MOLX activity. Subjects at risk for a disease that is caused or contributed to by aberrant MOLX expression or activity can be identified by, for example, any or a combination of diagnostic or prognostic assays as described herein. Administration of a prophylactic agent can occur prior to the manifestation of symptoms characteristic of the MOLX aberrancy, such that a disease or disorder is prevented or, alternatively, delayed in its progression. Depending upon the type of MOLX aberrancy, for example, an MOLX agonist or MOLX antagonist agent can be used for treating the subject. The appropriate agent can be determined based on screening assays described herein. The prophylactic methods of the invention are further discussed in the following subsections.

Therapeutic Methods

Another aspect of the invention pertains to methods of modulating MOLX expression or activity for therapeutic purposes. The modulatory method of the invention involves contacting a cell with an agent that modulates one or more of the activities of MOLX protein activity associated with the cell. An agent that modulates MOLX protein activity can be an agent as described herein, such as a nucleic acid or a protein, a naturally-occurring cognate ligand of an MOLX protein, a peptide, an MOLX peptidomimetic, or other small molecule. In one embodiment, the agent stimulates one or more MOLX protein activity. Examples of such stimulatory agents include active MOLX protein and a nucleic acid molecule encoding MOLX that has been introduced into the cell. In another embodiment, the agent inhibits one or more MOLX protein activity. Examples of such inhibitory agents include antisense MOLX nucleic acid molecules and anti-MOLX antibodies. These modulatory methods can be performed *in vitro* (*e.g.*, by culturing the cell with the agent) or, alternatively, *in vivo* (*e.g.*, by administering the agent to a subject). As such, the invention provides methods of treating an individual afflicted with a disease or disorder characterized by aberrant expression or activity of an MOLX protein or nucleic acid molecule. In one embodiment, the method involves administering an agent (*e.g.*, an agent identified by a screening assay described herein), or combination of agents that modulates (*e.g.*, up-regulates or down-regulates) MOLX expression or activity. In another embodiment, the method involves administering an MOLX protein or nucleic acid molecule as therapy to compensate for reduced or aberrant MOLX expression or activity.

Stimulation of MOLX activity is desirable in situations in which MOLX is abnormally downregulated and/or in which increased MOLX activity is likely to have a beneficial effect.

One example of such a situation is where a subject has a disorder characterized by aberrant cell proliferation and/or differentiation (*e.g.*, cancer or immune associated disorders). Another example of such a situation is where the subject has a gestational disease (*e.g.*, preclampsia).

Determination of the Biological Effect of the Therapeutic

In various embodiments of the invention, suitable *in vitro* or *in vivo* assays are performed to determine the effect of a specific Therapeutic and whether its administration is indicated for treatment of the affected tissue.

In various specific embodiments, *in vitro* assays may be performed with representative cells of the type(s) involved in the patient's disorder, to determine if a given Therapeutic exerts the desired effect upon the cell type(s). Compounds for use in therapy may be tested in suitable animal model systems including, but not limited to rats, mice, chicken, cows, monkeys, rabbits, and the like, prior to testing in human subjects. Similarly, for *in vivo* testing, any of the animal model system known in the art may be used prior to administration to human subjects.

Prophylactic and Therapeutic Uses of the Compositions of the Invention

The MOLX nucleic acids and proteins of the invention are useful in potential prophylactic and therapeutic applications implicated in a variety of disorders including, but not limited to: metabolic disorders, diabetes, obesity, infectious disease, anorexia, cancer-associated cancer, neurodegenerative disorders, Alzheimer's Disease, Parkinson's Disorder, immune disorders, hematopoietic disorders, and the various dyslipidemias, metabolic disturbances associated with obesity, the metabolic syndrome X and wasting disorders associated with chronic diseases and various cancers.

As an example, a cDNA encoding the MOLX protein of the invention may be useful in gene therapy, and the protein may be useful when administered to a subject in need thereof. By way of non-limiting example, the compositions of the invention will have efficacy for treatment of patients suffering from: metabolic disorders, diabetes, obesity, infectious disease, anorexia, cancer-associated cachexia, cancer, neurodegenerative disorders, Alzheimer's Disease, Parkinson's Disorder, immune disorders, hematopoietic disorders, and the various dyslipidemias.

Both the novel nucleic acid encoding the MOLX protein, and the MOLX protein of the invention, or fragments thereof, may also be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed. A further use could be as an anti-bacterial molecule (*i.e.*, some peptides have been found to possess anti-bacterial

properties). These materials are further useful in the generation of antibodies which immunospecifically-bind to the novel substances of the invention for use in therapeutic or diagnostic methods.

Examples

Example 1. Quantitative expression analysis of clones in various cells and tissues

The quantitative expression of various clones was assessed using microtiter plates containing RNA samples from a variety of normal and pathology-derived cells, cell lines and tissues using real time quantitative PCR (RTQ PCR; TAQMAN®). RTQ PCR was performed on a Perkin-Elmer Biosystems ABI PRISM® 7700 Sequence Detection System. Various collections of samples are assembled on the plates, and referred to as Panel 1 (containing cells and cell lines from normal and cancer sources), Panel 2 (containing samples derived from tissues, in particular from surgical samples, from normal and cancer sources), Panel 3 (containing samples derived from a wide variety of cancer sources), Panel 4 (containing cells and cell lines from normal cells and cells related to inflammatory conditions) and Panel CNSD.01 (containing samples from normal and diseased brains).

First, the RNA samples were normalized to constitutively expressed genes such as β -actin and GAPDH. RNA (~50 ng total or ~1 ng polyA+) was converted to cDNA using the TAQMAN® Reverse Transcription Reagents Kit (PE Biosystems, Foster City, CA; Catalog No. N808-0234) and random hexamers according to the manufacturer's protocol. Reactions were performed in 20 ul and incubated for 30 min. at 48°C. cDNA (5 ul) was then transferred to a separate plate for the TAQMAN® reaction using β-actin and GAPDH TAQMAN® Assay Reagents (PE Biosystems; Catalog Nos. 4310881E and 4310884E, respectively) and TAQMAN® universal PCR Master Mix (PE Biosystems; Catalog No. 4304447) according to the manufacturer's protocol. Reactions were performed in 25 ul using the following parameters: 2 min. at 50°C; 10 min. at 95°C; 15 sec. at 95°C/1 min. at 60°C (40 cycles). Results were recorded as CT values (cycle at which a given sample crosses a threshold level of fluorescence) using a log scale, with the difference in RNA concentration between a given sample and the sample with the lowest CT value being represented as 2 to the power of delta CT. The percent relative expression is then obtained by taking the reciprocal of this RNA difference and multiplying by 100. The average CT values obtained for ß-actin and GAPDH were used to normalize RNA samples. The RNA sample generating the highest CT value required no further diluting, while all other samples were diluted relative to this sample according to their β -actin /GAPDH average CT values.

Normalized RNA (5 ul) was converted to cDNA and analyzed via TAQMAN® using One Step RT-PCR Master Mix Reagents (PE Biosystems; Catalog No. 4309169) and gene-specific primers according to the manufacturer's instructions. Probes and primers were designed for each assay according to Perkin Elmer Biosystem's *Primer Express* Software package (version I for Apple Computer's Macintosh Power PC) or a similar algorithm using the target sequence as input. Default settings were used for reaction conditions and the following parameters were set before selecting primers: primer concentration = 250 nM, primer melting temperature (T_m) range = 58°-60° C, primer optimal Tm = 59° C, maximum primer difference = 2° C, probe does not have 5' G, probe T_m must be 10° C greater than primer T_m, amplicon size 75 bp to 100 bp. The probes and primers selected (see below) were synthesized by Synthegen (Houston, TX, USA). Probes were double purified by HPLC to remove uncoupled dye and evaluated by mass spectroscopy to verify coupling of reporter and quencher dyes to the 5' and 3' ends of the probe, respectively. Their final concentrations were: forward and reverse primers, 900 nM each, and probe, 200nM.

PCR conditions:

Normalized RNA from each tissue and each cell line was spotted in each well of a 96 well PCR plate (Perkin Elmer Biosystems). PCR cocktails including two probes (a probe specific for the target clone and another gene-specific probe multiplexed with the target probe) were set up using 1X TaqMan™ PCR Master Mix for the PE Biosystems 7700, with 5 mM MgCl2, dNTPs (dA, G, C, U at 1:1:1:2 ratios), 0.25 U/ml AmpliTaq Gold™ (PE Biosystems), and 0.4 U/□l RNase inhibitor, and 0.25 U/μl reverse transcriptase. Reverse transcription was performed at 48° C for 30 minutes followed by amplification/PCR cycles as follows: 95° C 10 min, then 40 cycles of 95° C for 15 seconds, 60° C for 1 minute.

In the results for Panel 1, the following abbreviations are used:

ca. = carcinoma,

* = established from metastasis,

met = metastasis,

s cell var= small cell variant,

non-s = non-sm =non-small,

squam = squamous,

pl. eff = pl effusion = pleural effusion,

glio = glioma,

astro = astrocytoma, and

neuro = neuroblastoma.

Panel 2

The plates for Panel 2 generally include 2 control wells and 94 test samples composed of RNA or cDNA isolated from human tissue procured by surgeons working in close cooperation with the National Cancer Institute's Cooperative Human Tissue Network (CHTN) or the National Disease Research Initiative (NDRI). The tissues are derived from human malignancies and in cases where indicated many malignant tissues have "matched margins" obtained from noncancerous tissue just adjacent to the tumor. These are termed normal adjacent tissues and are denoted "NAT" in the results below. The tumor tissue and the "matched margins" are evaluated by two independent pathologists (the surgical pathologists and again by a pathologists at NDRI or CHTN). This analysis provides a gross histopathological assessment of tumor differentiation grade. Moreover, most samples include the original surgical pathology report that provides information regarding the clinical stage of the patient. These matched margins are taken from the tissue surrounding (i.e. immediately proximal) to the zone of surgery (designated "NAT", for normal adjacent tissue, in Table RR). In addition, RNA and cDNA samples were obtained from various human tissues derived from autopsies performed on elderly people or sudden death victims (accidents, etc.). These tissues were ascertained to be free of disease and were purchased from various commercial sources such as Clontech (Palo Alto, CA), Research Genetics, and Invitrogen.

RNA integrity from all samples is controlled for quality by visual assessment of agarose gel electropherograms using 28S and 18S ribosomal RNA staining intensity ratio as a guide (2:1 to 2.5:1 28s:18s) and the absence of low molecular weight RNAs that would be indicative of degradation products. Samples are controlled against genomic DNA contamination by RTQ PCR reactions run in the absence of reverse transcriptase using probe and primer sets designed to amplify across the span of a single exon.

Panel 3D

The plates of Panel 3D are comprised of 94 cDNA samples and two control samples. Specifically, 92 of these samples are derived from cultured human cancer cell lines, 2 samples of human primary cerebellar tissue and 2 controls. The human cell lines are generally

obtained from ATCC (American Type Culture Collection, Manassas, VA), National Cancer Institute or the German tumor cell bank and fall into the following tissue groups: Squamous cell carcinoma of the tongue, breast cancer, prostate cancer, melanoma, epidermoid carcinoma, sarcomas, bladder carcinomas, pancreatic cancers, kidney cancers, leukemias/lymphomas, ovarian/uterine/cervical, gastric, colon, lung and CNS cancer cell lines. In addition, there are two independent samples of cerebellum. These cells are all cultured under standard recommended conditions and RNA extracted using the standard procedures. The cell lines in panel 3D and 1.3D are of the most common cell lines used in the scientific literature.

RNA integrity from all samples is controlled for quality by visual assessment of agarose gel electropherograms using 28S and 18S ribosomal RNA staining intensity ratio as a guide (2:1 to 2.5:1 28s:18s) and the absence of low molecular weight RNAs that would be indicative of degradation products. Samples are controlled against genomic DNA contamination by RTQ PCR reactions run in the absence of reverse transcriptase using probe and primer sets designed to amplify across the span of a single exon.

Panel 4

Panel 4 includes samples on a 96 well plate (2 control wells, 94 test samples) composed of RNA (Panel 4r) or cDNA (Panel 4d) isolated from various human cell lines or tissues related to inflammatory conditions. Total RNA from control normal tissues such as colon and lung (Stratagene ,La Jolla, CA) and thymus and kidney (Clontech) were employed. Total RNA from liver tissue from cirrhosis patients and kidney from lupus patients was obtained from BioChain (Biochain Institute, Inc., Hayward, CA). Intestinal tissue for RNA preparation from patients diagnosed as having Crohn's disease and ulcerative colitis was obtained from the National Disease Research Interchange (NDRI) (Philadelphia, PA).

Astrocytes, lung fibroblasts, dermal fibroblasts, coronary artery smooth muscle cells, small airway epithelium, bronchial epithelium, microvascular dermal endothelial cells, microvascular lung endothelial cells, human pulmonary aortic endothelial cells, human umbilical vein endothelial cells were all purchased from Clonetics (Walkersville, MD) and grown in the media supplied for these cell types by Clonetics. These primary cell types were activated with various cytokines or combinations of cytokines for 6 and/or 12-14 hours, as indicated. The following cytokines were used; IL-1 beta at approximately 1-5 ng/ml, TNF alpha at approximately 5-10 ng/ml, IFN gamma at approximately 20-50 ng/ml, IL-4 at

approximately 5-10 ng/ml, IL-9 at approximately 5-10 ng/ml, IL-13 at approximately 5-10 ng/ml. Endothelial cells were sometimes starved for various times by culture in the basal media from Clonetics with 0.1% serum.

Mononuclear cells were prepared from blood of employees at CuraGen Corporation, using Ficoll. LAK cells were prepared from these cells by culture in DMEM 5% FCS (Hyclone), 100 μM non essential amino acids (Gibco/Life Technologies, Rockville, MD), 1 mM sodium pyruvate (Gibco), mercaptoethanol 5.5 x 10^{-5} M (Gibco), and 10 mM Hepes (Gibco) and Interleukin 2 for 4-6 days. Cells were then either activated with 10-20 ng/ml PMA and 1-2 μg/ml ionomycin, IL-12 at 5-10 ng/ml, IFN gamma at 20-50 ng/ml and IL-18 at 5-10 ng/ml for 6 hours. In some cases, mononuclear cells were cultured for 4-5 days in DMEM 5% FCS (Hyclone), 100 μM non essential amino acids (Gibco), 1 mM sodium pyruvate (Gibco), mercaptoethanol 5.5 x 10^{-5} M (Gibco), and 10 mM Hepes (Gibco) with PHA (phytohemagglutinin) or PWM (pokeweed mitogen) at approximately 5 μg/ml. Samples were taken at 24, 48 and 72 hours for RNA preparation. MLR (mixed lymphocyte reaction) samples were obtained by taking blood from two donors, isolating the mononuclear cells using Ficoll and mixing the isolated mononuclear cells 1:1 at a final concentration of approximately $2x10^6$ cells/ml in DMEM 5% FCS (Hyclone), 100 μM non essential amino acids (Gibco), 1 mM sodium pyruvate (Gibco), mercaptoethanol (5.5 x 10^{-5} M) (Gibco), and 10 mM Hepes (Gibco). The MLR was cultured and samples taken at various time points ranging from 1- 7 days for RNA preparation.

Monocytes were isolated from mononuclear cells using CD14 Miltenyi Beads, +ve VS selection columns and a Vario Magnet according to the manufacturer's instructions. Monocytes were differentiated into dendritic cells by culture in DMEM 5% fetal calf serum (FCS) (Hyclone, Logan, UT), 100 μM non essential amino acids (Gibco), 1 mM sodium pyruvate (Gibco), mercaptoethanol 5.5 x 10^{-5} M (Gibco), and 10 mM Hepes (Gibco), 50 ng/ml GMCSF and 5 ng/ml IL-4 for 5-7 days. Macrophages were prepared by culture of monocytes for 5-7 days in DMEM 5% FCS (Hyclone), 100 μM non essential amino acids (Gibco), 1 mM sodium pyruvate (Gibco), mercaptoethanol 5.5 x 10^{-5} M (Gibco), 10 mM Hepes (Gibco) and 10% AB Human Serum or MCSF at approximately 50 ng/ml. Monocytes, macrophages and dendritic cells were stimulated for 6 and 12-14 hours with lipopolysaccharide (LPS) at 100 ng/ml. Dendritic cells were also stimulated with anti-CD40 monoclonal antibody (Pharmingen) at 10 μg/ml for 6 and 12-14 hours.

CD4 lymphocytes, CD8 lymphocytes and NK cells were also isolated from mononuclear cells using CD4, CD8 and CD56 Miltenyi beads, positive VS selection columns

and a Vario Magnet according to the manufacturer's instructions. CD45RA and CD45RO CD4 lymphocytes were isolated by depleting mononuclear cells of CD8, CD56, CD14 and CD19 cells using CD8, CD56, CD14 and CD19 Miltenyi beads and positive selection. Then CD45RO beads were used to isolate the CD45RO CD4 lymphocytes with the remaining cells being CD45RA CD4 lymphocytes. CD45RA CD4, CD45RO CD4 and CD8 lymphocytes were placed in DMEM 5% FCS (Hyclone), 100 μM non essential amino acids (Gibco), 1 mM sodium pyruvate (Gibco), mercaptoethanol 5.5 x 10^{-5} M (Gibco), and 10 mM Hepes (Gibco) and plated at 10^6 cells/ml onto Falcon 6 well tissue culture plates that had been coated overnight with 0.5 μg/ml anti-CD28 (Pharmingen) and 3 ug/ml anti-CD3 (OKT3, ATCC) in PBS. After 6 and 24 hours, the cells were harvested for RNA preparation. To prepare chronically activated CD8 lymphocytes, we activated the isolated CD8 lymphocytes for 4 days on anti-CD28 and anti-CD3 coated plates and then harvested the cells and expanded them in DMEM 5% FCS (Hyclone), 100 μM non essential amino acids (Gibco), 1 mM sodium pyruvate (Gibco), mercaptoethanol 5.5 x 10^{-5} M (Gibco), and 10 mM Hepes (Gibco) and IL-2. The expanded CD8 cells were then activated again with plate bound anti-CD3 and anti-CD28 for 4 days and expanded as before. RNA was isolated 6 and 24 hours after the second activation and after 4 days of the second expansion culture. The isolated NK cells were cultured in DMEM 5% FCS (Hyclone), 100 μM non essential amino acids (Gibco), 1 mM sodium pyruvate (Gibco), mercaptoethanol 5.5 x 10^{-5} M (Gibco), and 10 mM Hepes (Gibco) and IL-2 for 4-6 days before RNA was prepared.

To obtain B cells, tonsils were procured from NDRI. The tonsil was cut up with sterile dissecting scissors and then passed through a sieve. Tonsil cells were then spun down and resupended at 10^6 cells/ml in DMEM 5% FCS (Hyclone), 100 μM non essential amino acids (Gibco), 1 mM sodium pyruvate (Gibco), mercaptoethanol 5.5 x 10^{-5} M (Gibco), and 10 mM Hepes (Gibco). To activate the cells, we used PWM at 5 μg/ml or anti-CD40 (Pharmingen) at approximately 10 μg/ml and IL-4 at 5-10 ng/ml. Cells were harvested for RNA preparation at 24,48 and 72 hours.

To prepare the primary and secondary Th1/Th2 and Tr1 cells, six-well Falcon plates were coated overnight with 10 μg/ml anti-CD28 (Pharmingen) and 2 μg/ml OKT3 (ATCC), and then washed twice with PBS. Umbilical cord blood CD4 lymphocytes (Poietic Systems, German Town, MD) were cultured at 10^5-10^6 cells/ml in DMEM 5% FCS (Hyclone), 100 μM non essential amino acids (Gibco), 1 mM sodium pyruvate (Gibco), mercaptoethanol 5.5 x 10^{-5} M (Gibco), 10 mM Hepes (Gibco) and IL-2 (4 ng/ml). IL-12 (5 ng/ml) and anti-IL4 (1 □g/ml) were used to direct to Th1, while IL-4 (5 ng/ml) and anti-IFN gamma (1 □g/ml) were

used to direct to Th2 and IL-10 at 5 ng/ml was used to direct to Tr1. After 4-5 days, the activated Th1, Th2 and Tr1 lymphocytes were washed once in DMEM and expanded for 4-7 days in DMEM 5% FCS (Hyclone), 100 μM non essential amino acids (Gibco), 1 mM sodium pyruvate (Gibco), mercaptoethanol 5.5 x 10^{-5} M (Gibco), 10 mM Hepes (Gibco) and IL-2 (1 ng/ml). Following this, the activated Th1, Th2 and Tr1 lymphocytes were re-stimulated for 5 days with anti-CD28/OKT3 and cytokines as described above, but with the addition of anti-CD95L (1 □g/ml) to prevent apoptosis. After 4-5 days, the Th1, Th2 and Tr1 lymphocytes were washed and then expanded again with IL-2 for 4-7 days. Activated Th1 and Th2 lymphocytes were maintained in this way for a maximum of three cycles. RNA was prepared from primary and secondary Th1, Th2 and Tr1 after 6 and 24 hours following the second and third activations with plate bound anti-CD3 and anti-CD28 mAbs and 4 days into the second and third expansion cultures in Interleukin 2.

The following leukocyte cells lines were obtained from the ATCC: Ramos, EOL-1, KU-812. EOL cells were further differentiated by culture in 0.1 mM dbcAMP at 5 x10^5 cells/ml for 8 days, changing the media every 3 days and adjusting the cell concentration to 5 x10^5 cells/ml. For the culture of these cells, we used DMEM or RPMI (as recommended by the ATCC), with the addition of 5% FCS (Hyclone), 100 μM non essential amino acids (Gibco), 1 mM sodium pyruvate (Gibco), mercaptoethanol 5.5 x 10^{-5} M (Gibco), 10 mM Hepes (Gibco). RNA was either prepared from resting cells or cells activated with PMA at 10 ng/ml and ionomycin at 1 μg/ml for 6 and 14 hours. Keratinocyte line CCD106 and an airway epithelial tumor line NCI-H292 were also obtained from the ATCC. Both were cultured in DMEM 5% FCS (Hyclone), 100 μM non essential amino acids (Gibco), 1 mM sodium pyruvate (Gibco), mercaptoethanol 5.5 x 10^{-5} M (Gibco), and 10 mM Hepes (Gibco). CCD1106 cells were activated for 6 and 14 hours with approximately 5 ng/ml TNF alpha and 1 ng/ml IL-1 beta, while NCI-H292 cells were activated for 6 and 14 hours with the following cytokines: 5 ng/ml IL-4, 5 ng/ml IL-9, 5 ng/ml IL-13 and 25 ng/ml IFN gamma.

For these cell lines and blood cells, RNA was prepared by lysing approximately 10^7 cells/ml using Trizol (Gibco BRL). Briefly, 1/10 volume of bromochloropropane (Molecular Research Corporation) was added to the RNA sample, vortexed and after 10 minutes at room temperature, the tubes were spun at 14,000 rpm in a Sorvall SS34 rotor. The aqueous phase was removed and placed in a 15 ml Falcon Tube. An equal volume of isopropanol was added and left at –20 degrees C overnight. The precipitated RNA was spun down at 9,000 rpm for 15 min in a Sorvall SS34 rotor and washed in 70% ethanol. The pellet was redissolved in 300 μl of RNAse-free water and 35 μl buffer (Promega) 5 μl DTT, 7 μl

RNAsin and 8 µl DNAse were added. The tube was incubated at 37 degrees C for 30 minutes to remove contaminating genomic DNA, extracted once with phenol chloroform and re-precipitated with 1/10 volume of 3 M sodium acetate and 2 volumes of 100% ethanol. The RNA was spun down and placed in RNAse free water. RNA was stored at –80 degrees C.

Panel CNSD.01

The plates for Panel CNSD.01 include two control wells and 94 test samples comprised of cDNA isolated from postmortem human brain tissue obtained from the Harvard Brain Tissue Resource Center. Brains are removed from calvaria of donors between 4 and 24 hours after death, sectioned by neuroanatomists, and frozen at -80°C in liquid nitrogen vapor. All brains are sectioned and examined by neuropathologists to confirm diagnoses with clear associated neuropathology.

Disease diagnoses are taken from patient records. The panel contains two brains from each of the following diagnoses: Alzheimer's disease, Parkinson's disease, Huntington's disease, Progressive Supernuclear Palsy, Depression, and "Normal controls". Within each of these brains, the following regions are represented: cingulate gyrus, temporal pole, globus palladus, substantia nigra, Brodman Area 4 (primary motor strip), Brodman Area 7 (parietal cortex), Brodman Area 9 (prefrontal cortex), and Brodman area 17 (occipital cortex). Not all brain regions are represented in all cases; e.g., Huntington's disease is characterized in part by neurodegeneration in the globus palladus, thus this region is impossible to obtain from confirmed Huntington's cases. Likewise Parkinson's disease is characterized by degeneration of the substantia nigra making this region more difficult to obtain. Normal control brains were examined for neuropathology and found to be free of any pathology consistent with neurodegeneration.

RNA integrity from all samples is controlled for quality by visual assessment of agarose gel electropherograms using 28S and 18S ribosomal RNA staining intensity ratio as a guide (2:1 to 2.5:1 28s:18s) and the absence of low molecular weight RNAs that would be indicative of degradation products. Samples are controlled against genomic DNA contamination by RTQ PCR reactions run in the absence of reverse transcriptase using probe and primer sets designed to amplify across the span of a single exon.

In the labels employed to identify tissues in the CNS panel, the following abbreviations are used:

PSP = Progressive supranuclear palsy

Sub Nigra = Substantia nigra

Glob Palladus= Globus palladus

Temp Pole = Temporal pole

Cing Gyr = Cingulate gyrus

BA 4 = Brodman Area 4

A. MOL1a

Expression of gene SC29674552_EXT was assessed using the primer-probe sets Ag267 and Ag1308, described in Tables 10 and 11. Results of the RTQ-PCR runs are shown in Tables 12, 13, 14, 15, and 16

Table 10. Probe Name Ag267

Primers	Sequences	TM	Length	Start Position	SEQ ID NO:
Forward	5'-TGCAGCGACCATCGTTCA-3'		18	50	65
Probe	TET-5'-CTGCTGTAACATTCATCAATCTGGTCACTGCA-3'-TAMRA		32	76	66
Reverse	5'-GGGTACATGGGCGCCAT-3'		17	109	67

Table 11. Probe Name: Ag1308

Primers	Sequences	TM	Length	Start Position	SEQ ID NO:
Forward	5'-GAGTGTGACATTCCAGGACACT-3'	59.1	22	139	68
Probe	FAM-5'-ATGGTGGCACCTGCCTCAACCTG-3'-TAMRA	71.6	23	167	69
Reverse	5'-GCACTGGCACTGGTAGGAA-3'	59.8	19	195	70

Table 12. Panel 1

Tissue Name	Relative Expression(%) 1tm420t_ag267	Tissue Name	Relative Expression(%) 1tm420t_ag267
Endothelial cells	4.4	Kidney (fetal)	19.8
Endothelial cells (treated)	5.3	Renal ca. 786-0	17.4

Pancreas	11.5	Renal ca. A498	20.6
Pancreatic ca. CAPAN 2	7.0	Renal ca. RXF 393	18.4
Adipose	40.3	Renal ca. ACHN	20.4
Adrenal gland	5.8	Renal ca. UO-31	20.3
Thyroid	11.5	Renal ca. TK-10	**100.0**
Salivary gland	7.5	Liver	7.3
Pituitary gland	4.9	Liver (fetal)	2.3
Brain (fetal)	1.2	Liver ca. (hepatoblast) HepG2	0.0
Brain (whole)	7.9	Lung	5.4
Brain (amygdala)	5.3	Lung (fetal)	4.2
Brain (cerebellum)	16.6	Lung ca. (small cell) LX-1	2.8
Brain (hippocampus)	1.8	Lung ca. (small cell) NCI-H69	5.5
Brain (substantia nigra)	2.0	Lung ca. (s.cell var.) SHP-77	3.2
Brain (thalamus)	1.6	Lung ca. (large cell)NCI-H460	6.7
Brain (hypothalamus)	0.9	Lung ca. (non-sm. cell) A549	6.5
Spinal cord	3.0	Lung ca. (non-s.cell) NCI-H23	9.0
CNS ca. (glio/astro) U87-MG	70.7	Lung ca (non-s.cell) HOP-62	14.4
CNS ca. (glio/astro) U-118-MG	21.6	Lung ca. (non-s.cl) NCI-H522	56.6
CNS ca. (astro) SW1783	20.7	Lung ca. (squam.) SW 900	28.1
CNS ca.* (neuro; met) SK-N-AS	0.0	Lung ca. (squam.) NCI-H596	2.9
CNS ca. (astro) SF-539	14.6	Mammary gland	35.1
CNS ca. (astro) SNB-75	15.8	Breast ca.* (pl. effusion) MCF-7	32.3
CNS ca. (glio) SNB-19	62.8	Breast ca.* (pl.ef) MDA-MB-231	19.6
CNS ca. (glio) U251	5.9	Breast ca.* (pl. effusion) T47D	19.1
CNS ca. (glio) SF-295	23.0	Breast ca. BT-549	11.0
Heart	5.7	Breast ca. MDA-N	7.9
Skeletal muscle	1.3	Ovary	44.4
Bone marrow	4.1	Ovarian ca. OVCAR-3	15.9
Thymus	32.3	Ovarian ca. OVCAR-4	44.4
Spleen	3.4	Ovarian ca. OVCAR-5	30.1
Lymph node	5.0	Ovarian ca. OVCAR-8	50.7
Colon (ascending)	16.8	Ovarian ca. IGROV-1	10.4
Stomach	11.1	Ovarian ca.* (ascites) SK-OV-3	11.9
Small intestine	2.6	Uterus	8.3
Colon ca. SW480	6.2	Placenta	21.0
Colon ca.* (SW480 met)SW620	0.9	Prostate	11.3
Colon ca. HT29	6.2	Prostate ca.* (bone met)PC-3	22.2
Colon ca. HCT-116	19.2	Testis	92.0
Colon ca. CaCo-2	15.0	Melanoma Hs688(A).T	13.5
Colon ca. HCT-15	20.2	Melanoma* (met) Hs688(B).T	20.3
Colon ca. HCC-2998	3.9	Melanoma UACC-62	8.5
Gastric ca.* (liver met) NCI-N87	22.8	Melanoma M14	15.2

Bladder	13.3	Melanoma LOX IMVI	54.7
Trachea	10.3	Melanoma* (met) SK-MEL-5	6.8
Kidney	6.4	Melanoma SK-MEL-28	38.4

Table 13. Panel 1.2

Tissue Name	Relative Expression(%) 1.2tm1287t_ag267	Tissue Name	Relative Expression(%) 1.2tm1287t_ag267
Endothelial cells	8.2	Renal ca. 786-0	16.8
Endothelial cells (treated)	15.0	Renal ca. A498	32.5
Pancreas	1.4	Renal ca. RXF 393	15.6
Pancreatic ca. CAPAN 2	2.9	Renal ca. ACHN	15.3
Adrenal Gland (new lot*)	14.7	Renal ca. UO-31	15.9
Thyroid	2.1	Renal ca. TK-10	53.2
Salivary gland	14.8	Liver	7.7
Pituitary gland	3.2	Liver (fetal)	3.8
Brain (fetal)	1.9	Liver ca. (hepatoblast) HepG2	0.0
Brain (whole)	7.1	Lung	6.0
Brain (amygdala)	7.7	Lung (fetal)	3.6
Brain (cerebellum)	5.0	Lung ca. (small cell) LX-1	2.0
Brain (hippocampus)	8.7	Lung ca. (small cell) NCI-H69	4.5
Brain (thalamus)	1.5	Lung ca. (s.cell var.) SHP-77	0.3
Cerebral Cortex	52.8	Lung ca. (large cell)NCI-H460	14.8
Spinal cord	6.4	Lung ca. (non-sm. cell) A549	8.3
CNS ca. (glio/astro) U87-MG	100.0	Lung ca. (non-s.cell) NCI-H23	13.4
CNS ca. (glio/astro) U-118-MG	37.1	Lung ca (non-s.cell) HOP-62	30.4
CNS ca. (astro) SW1783	14.2	Lung ca. (non-s.cl) NCI-H522	57.4
CNS ca.* (neuro; met) SK-N-AS	0.1	Lung ca. (squam.) SW 900	20.9
CNS ca. (astro) SF-539	11.9	Lung ca. (squam.) NCI-H596	2.7
CNS ca. (astro) SNB-75	9.6	Mammary gland	13.3
CNS ca. (glio) SNB-19	80.7	Breast ca.* (pl. effusion) MCF-7	31.9
CNS ca. (glio) U251	26.6	Breast ca.* (pl.ef) MDA-MB-231	23.7
CNS ca. (glio) SF-295	6.6	Breast ca.* (pl. effusion) T47D	12.2
Heart	21.0	Breast ca. BT-549	11.7
Skeletal Muscle (new lot*)	4.9	Breast ca. MDA-N	12.2
Bone marrow	2.2	Ovary	34.2
Thymus	2.1	Ovarian ca. OVCAR-3	35.8
Spleen	2.7	Ovarian ca. OVCAR-4	32.3
Lymph node	9.9	Ovarian ca. OVCAR-5	21.0
Colorectal	8.8	Ovarian ca. OVCAR-8	32.5
Stomach	8.7	Ovarian ca. IGROV-1	20.0

Small intestine	4.2	Ovarian ca.* (ascites) SK-OV-3	22.2
Colon ca. SW480	2.5	Uterus	5.2
Colon ca.* (SW480 met)SW620	0.5	Placenta	24.3
Colon ca. HT29	1.4	Prostate	9.2
Colon ca. HCT-116	13.4	Prostate ca.* (bone met)PC-3	30.6
Colon ca. CaCo-2	7.5	Testis	8.7
83219 CC Well to Mod Diff (ODO3866)	6.9	Melanoma Hs688(A).T	10.4
Colon ca. HCC-2998	14.7	Melanoma* (met) Hs688(B).T	15.8
Gastric ca.* (liver met) NCI-N87	38.2	Melanoma UACC-62	27.0
Bladder	37.1	Melanoma M14	13.1
Trachea	3.5	Melanoma LOX IMVI	8.4
Kidney	7.3	Melanoma* (met) SK-MEL-5	12.9
Kidney (fetal)	29.5	Adipose	27.5

Table 14. Panel 2D

Tissue Name	Relative Expression(%) 2Dtm2336t_ ag267	Tissue Name	Relative Expression(%) 2Dtm2336t_ ag267
Normal Colon GENPAK 061003	23.7	Kidney NAT Clontech 8120608	7.7
83219 CC Well to Mod Diff (ODO3866)	6.1	Kidney Cancer Clontech 8120613	0.7
83220 CC NAT (ODO3866)	8.2	Kidney NAT Clontech 8120614	10.2
83221 CC Gr.2 rectosigmoid (ODO3868)	3.8	Kidney Cancer Clontech 9010320	24.0
83222 CC NAT (ODO3868)	3.5	Kidney NAT Clontech 9010321	24.3
83235 CC Mod Diff (ODO3920)	3.7	Normal Uterus GENPAK 061018	8.4
83236 CC NAT (ODO3920)	6.4	Uterus Cancer GENPAK 064011	15.5
83237 CC Gr.2 ascend colon (ODO3921)	7.5	Normal Thyroid Clontech A+ 6570-1	11.0
83238 CC NAT (ODO3921)	4.0	Thyroid Cancer GENPAK 064010	27.9
83241 CC from Partial Hepatectomy (ODO4309)	7.6	Thyroid Cancer INVITROGEN A302152	14.0
83242 Liver NAT (ODO4309)	7.5	Thyroid NAT INVITROGEN A302153	21.6
87472 Colon mets to lung (OD04451-01)	6.0	Normal Breast GENPAK 061019	33.7
87473 Lung NAT (OD04451-02)	8.4	84877 Breast Cancer (OD04566)	11.9
Normal Prostate Clontech A+	7.7	85975 Breast Cancer	18.3

6546-1		(OD04590-01)	
84140 Prostate Cancer (OD04410)	22.1	85976 Breast Cancer Mets (OD04590-03)	37.1
84141 Prostate NAT (OD04410)	19.8	87070 Breast Cancer Metastasis (OD04655-05)	31.4
87073 Prostate Cancer (OD04720-01)	12.5	GENPAK Breast Cancer 064006	15.2
87074 Prostate NAT (OD04720-02)	27.5	Breast Cancer Res. Gen. 1024	30.8
Normal Lung GENPAK 061010	21.5	Breast Cancer Clontech 9100266	100.0
83239 Lung Met to Muscle (ODO4286)	10.2	Breast NAT Clontech 9100265	45.4
83240 Muscle NAT (ODO4286)	10.2	Breast Cancer INVITROGEN A209073	29.3
84136 Lung Malignant Cancer (OD03126)	16.0	Breast NAT INVITROGEN A2090734	26.1
84137 Lung NAT (OD03126)	15.9	Normal Liver GENPAK 061009	6.1
84871 Lung Cancer (OD04404)	8.0	Liver Cancer GENPAK 064003	11.7
84872 Lung NAT (OD04404)	19.2	Liver Cancer Research Genetics RNA 1025	6.4
84875 Lung Cancer (OD04565)	3.4	Liver Cancer Research Genetics RNA 1026	10.6
84876 Lung NAT (OD04565)	8.4	Paired Liver Cancer Tissue Research Genetics RNA 6004-T	14.7
85950 Lung Cancer (OD04237-01)	27.4	Paired Liver Tissue Research Genetics RNA 6004-N	4.4
85970 Lung NAT (OD04237-02)	15.8	Paired Liver Cancer Tissue Research Genetics RNA 6005-T	11.5
83255 Ocular Mel Met to Liver (ODO4310)	11.0	Paired Liver Tissue Research Genetics RNA 6005-N	5.4
83256 Liver NAT (ODO4310)	7.0	Normal Bladder GENPAK 061001	19.5
84139 Melanoma Mets to Lung (OD04321)	12.1	Bladder Cancer Research Genetics RNA 1023	9.3
84138 Lung NAT (OD04321)	23.5	Bladder Cancer INVITROGEN A302173	14.7
Normal Kidney GENPAK 061008	25.2	87071 Bladder Cancer (OD04718-01)	13.3
83786 Kidney Ca, Nuclear grade 2 (OD04338)	24.1	87072 Bladder Normal Adjacent (OD04718-03)	17.3
83787 Kidney NAT (OD04338)	7.7	Normal Ovary Res. Gen.	14.5
83788 Kidney Ca Nuclear grade 1/2 (OD04339)	14.5	Ovarian Cancer GENPAK 064008	26.8
83789 Kidney NAT (OD04339)	9.2	87492 Ovary Cancer (OD04768-07)	15.5
83790 Kidney Ca, Clear cell type (OD04340)	19.9	87493 Ovary NAT (OD04768-08)	9.9
83791 Kidney NAT (OD04340)	15.3	Normal Stomach GENPAK 061017	6.3

83792 Kidney Ca, Nuclear grade 3 (OD04348)	14.6	Gastric Cancer Clontech 9060358	3.7
83793 Kidney NAT (OD04348)	14.3	NAT Stomach Clontech 9060359	5.0
87474 Kidney Cancer (OD04622-01)	19.3	Gastric Cancer Clontech 9060395	11.8
87475 Kidney NAT (OD04622-03)	5.3	NAT Stomach Clontech 9060394	5.9
85973 Kidney Cancer (OD04450-01)	17.2	Gastric Cancer Clontech 9060397	14.5
85974 Kidney NAT (OD04450-03)	10.4	NAT Stomach Clontech 9060396	2.8
Kidney Cancer Clontech 8120607	17.9	Gastric Cancer GENPAK 064005	5.7

Table 15. Panel 2.2

Tissue Name	Relative Expression(%) 2.2x4tm6515f_ag1308_b2	Tissue Name	Relative Expression(%) 2.2x4tm6515f_ag1308_b2
Normal Colon GENPAK 061003	26.3	83793 Kidney NAT (OD04348)	83.7
97759 Colon cancer (OD06064)	20.5	98938 Kidney malignant cancer (OD06204B)	9.7
97760 Colon cancer NAT (OD06064)	6.5	98939 Kidney normal adjacent tissue (OD06204E)	15.3
97778 Colon cancer (OD06159)	2.6	85973 Kidney Cancer (OD04450-01)	55.1
97779 Colon cancer NAT (OD06159)	17.2	85974 Kidney NAT (OD04450-03)	23.4
98861 Colon cancer (OD06297-04)	2.7	Kidney Cancer Clontech 8120613	2.1
98862 Colon cancer NAT (OD06297-015)	25.9	Kidney NAT Clontech 8120614	12.5
83237 CC Gr.2 ascend colon (ODO3921)	9.2	Kidney Cancer Clontech 9010320	12.6
83238 CC NAT (ODO3921)	11.3	Kidney NAT Clontech 9010321	7.0
97766 Colon cancer metastasis (OD06104)	5.4	Kidney Cancer Clontech 8120607	19.0
97767 Lung NAT (OD06104)	4.6	Kidney NAT Clontech 8120608	6.5
87472 Colon mets to lung (OD04451-01)	22.6	Normal Uterus GENPAK 061018	31.5
87473 Lung NAT (OD04451-02)	18.3	Uterus Cancer GENPAK 064011	21.4
Normal Prostate Clontech A+ 6546-1 (8090438)	7.2	Normal Thyroid Clontech A+ 6570-1 (7080817)	1.2
84140 Prostate Cancer (OD04410)	11.0	Thyroid Cancer GENPAK 064010	16.7
84141 Prostate NAT (OD04410)	20.3	Thyroid Cancer INVITROGEN A302152	44.5

Normal Ovary Res. Gen.	29.1	Thyroid NAT INVITROGEN A302153	14.0
98863 Ovarian cancer (OD06283-03)	31.5	Normal Breast GENPAK 061019	39.9
98865 Ovarian cancer NAT/fallopian tube (OD06283-07)	19.3	84877 Breast Cancer (OD04566)	11.0
Ovarian Cancer GENPAK 064008	22.2	Breast Cancer Res. Gen. 1024	49.2
97773 Ovarian cancer (OD06145)	12.4	85975 Breast Cancer (OD04590-01)	50.8
97775 Ovarian cancer NAT (OD06145)	36.2	85976 Breast Cancer Mets (OD04590-03)	42.9
98853 Ovarian cancer (OD06455-03)	9.3	87070 Breast Cancer Metastasis (OD04655-05)	72.1
98854 Ovarian NAT (OD06455-07) Fallopian tube	7.6	GENPAK Breast Cancer 064006	29.0
Normal Lung GENPAK 061010	25.4	Breast Cancer Clontech 9100266	41.1
92337 Invasive poor diff. lung adeno (ODO4945-01	10.9	Breast NAT Clontech 9100265	15.6
92338 Lung NAT (ODO4945-03)	20.2	Breast Cancer INVITROGEN A209073	13.5
84136 Lung Malignant Cancer (OD03126)	10.5	Breast NAT INVITROGEN A2090734	43.8
84137 Lung NAT (OD03126)	24.5	97763 Breast cancer (OD06083)	49.6
90372 Lung Cancer (OD05014A)	25.9	97764 Breast cancer node metastasis (OD06083)	44.3
90373 Lung NAT (OD05014B)	26.1	Normal Liver GENPAK 061009	38.6
97761 Lung cancer (OD06081)	8.3	Liver Cancer Research Genetics RNA 1026	13.2
97762 Lung cancer NAT (OD06081)	19.4	Liver Cancer Research Genetics RNA 1025	36.4
85950 Lung Cancer (OD04237-01)	13.3	Paired Liver Cancer Tissue Research Genetics RNA 6004-T	23.8
85970 Lung NAT (OD04237-02)	41.6	Paired Liver Tissue Research Genetics RNA 6004-N	5.9
83255 Ocular Mel Met to Liver (ODO4310)	12.9	Paired Liver Cancer Tissue Research Genetics RNA 6005-T	25.1
83256 Liver NAT (ODO4310)	11.5	Paired Liver Tissue Research Genetics RNA 6005-N	47.4
84139 Melanoma Mets to Lung (OD04321)	22.9	Liver Cancer GENPAK 064003	36.9
84138 Lung NAT (OD04321)	13.6	Normal Bladder GENPAK 061001	18.9
Normal Kidney GENPAK 061008	19.9	Bladder Cancer Research Genetics RNA 1023	9.6
83786 Kidney Ca, Nuclear grade 2 (OD04338)	59.8	Bladder Cancer INVITROGEN A302173	24.5

83787 Kidney NAT (OD04338)	23.8	Normal Stomach GENPAK 061017	43.1
83788 Kidney Ca Nuclear grade 1/2 (OD04339)	100.0	Gastric Cancer Clontech 9060397	6.2
83789 Kidney NAT (OD04339)	7.3	NAT Stomach Clontech 9060396	7.3
83790 Kidney Ca, Clear cell type (OD04340)	15.0	Gastric Cancer Clontech 9060395	9.3
83791 Kidney NAT (OD04340)	22.5	NAT Stomach Clontech 9060394	13.3
83792 Kidney Ca, Nuclear grade 3 (OD04348)	14.3	Gastric Cancer GENPAK 064005	9.4

Table 16. Panel 4D

Tissue Name	Relative Expression(%) 4Dtm1935t_ag267	Relative Expression(%) 4Dtm1888f_ag1308
93768_Secondary Th1_anti-CD28/anti-CD3	51.0	23.2
93769_Secondary Th2_anti-CD28/anti-CD3	43.5	24.1
93770_Secondary Tr1_anti-CD28/anti-CD3	42.6	23.8
93573_Secondary Th1_resting day 4-6 in IL-2	11.7	7.4
93572_Secondary Th2_resting day 4-6 in IL-2	21.9	12.4
93571_Secondary Tr1_resting day 4-6 in IL-2	12.1	8.0
93568_primary Th1_anti-CD28/anti-CD3	47.0	29.5
93569_primary Th2_anti-CD28/anti-CD3	28.1	22.8
93570_primary Tr1_anti-CD28/anti-CD3	45.1	37.9
93565_primary Th1_resting dy 4-6 in IL-2	51.8	49.3
93566_primary Th2_resting dy 4-6 in IL-2	23.0	27.9
93567_primary Tr1_resting dy 4-6 in IL-2	34.2	27.4
93351_CD45RA CD4 lymphocyte_anti-CD28/anti-CD3	32.3	16.5
93352_CD45RO CD4 lymphocyte_anti-CD28/anti-CD3	52.5	29.3
93251_CD8 Lymphocytes_anti-CD28/anti-CD3	22.8	14.9
93353_chronic CD8 Lymphocytes 2ry_resting dy 4-6 in IL-2	20.2	12.6
93574_chronic CD8 Lymphocytes 2ry_activated CD3/CD28	22.1	14.2
93354_CD4_none	6.8	8.0
93252_Secondary Th1/Th2/Tr1_anti-CD95 CH11	31.0	23.0
93103_LAK cells_resting	62.0	30.8
93788_LAK cells_IL-2	47.3	27.4
93787_LAK cells_IL-2+IL-12	41.5	23.3
93789_LAK cells_IL-2+IFN gamma	36.1	27.4
93790_LAK cells_IL-2+ IL-18	27.0	19.8
93104_LAK cells_PMA/ionomycin and IL-18	38.2	21.2
93578_NK Cells IL-2_resting	17.0	14.7
93109_Mixed Lymphocyte Reaction_Two Way MLR	30.8	27.7
93110_Mixed Lymphocyte Reaction_Two Way MLR	19.6	12.2

93111_Mixed Lymphocyte Reaction_Two Way MLR	17.3	10.4
93112_Mononuclear Cells (PBMCs)_resting	23.0	17.2
93113_Mononuclear Cells (PBMCs)_PWM	91.4	56.6
93114_Mononuclear Cells (PBMCs)_PHA-L	52.5	31.9
93249_Ramos (B cell)_none	14.6	14.7
93250_Ramos (B cell)_ionomycin	18.2	21.9
93349_B lymphocytes_PWM	43.2	30.6
93350_B lymphoytes_CD40L and IL-4	12.7	11.3
92665_EOL-1 (Eosinophil)_dbcAMP differentiated	5.4	5.7
93248_EOL-1 (Eosinophil)_dbcAMP/PMAionomycin	14.2	12.9
93356_Dendritic Cells_none	28.5	16.2
93355_Dendritic Cells_LPS 100 ng/ml	23.0	14.7
93775_Dendritic Cells_anti-CD40	21.5	12.3
93774_Monocytes_resting	81.8	58.2
93776_Monocytes_LPS 50 ng/ml	**100.0**	**100.0**
93581_Macrophages_resting	75.3	33.4
93582_Macrophages_LPS 100 ng/ml	54.3	30.8
93098_HUVEC (Endothelial)_none	12.7	6.3
93099_HUVEC (Endothelial)_starved	18.3	12.7
93100_HUVEC (Endothelial)_IL-1b	6.9	4.9
93779_HUVEC (Endothelial)_IFN gamma	11.7	6.9
93102_HUVEC (Endothelial)_TNF alpha + IFN gamma	12.9	6.4
93101_HUVEC (Endothelial)_TNF alpha + IL4	20.7	10.7
93781_HUVEC (Endothelial)_IL-11	4.8	2.8
93583_Lung Microvascular Endothelial Cells_none	14.8	7.4
93584_Lung Microvascular Endothelial Cells_TNFa (4 ng/ml) and IL1b (1 ng/ml)	25.3	9.8
92662_Microvascular Dermal endothelium_none	27.4	16.4
92663_Microsvasular Dermal endothelium_TNFa (4 ng/ml) and IL1b (1 ng/ml)	40.3	17.8
93773_Bronchial epithelium_TNFa (4 ng/ml) and IL1b (1 ng/ml) **	37.6	24.1
93347_Small Airway Epithelium_none	11.6	7.3
93348_Small Airway Epithelium_TNFa (4 ng/ml) and IL1b (1 ng/ml)	81.8	53.2
92668_Coronery Artery SMC_resting	38.2	24.5
92669_Coronery Artery SMC_TNFa (4 ng/ml) and IL1b (1 ng/ml)	32.3	24.0
93107_astrocytes_resting	14.7	8.6
93108_astrocytes_TNFa (4 ng/ml) and IL1b (1 ng/ml)	24.7	14.2
92666_KU-812 (Basophil)_resting	12.4	9.9
92667_KU-812 (Basophil)_PMA/ionoycin	46.7	29.3
93579_CCD1106 (Keratinocytes)_none	13.7	8.9
93580_CCD1106 (Keratinocytes)_TNFa and IFNg **	81.8	57.4
93791_Liver Cirrhosis	5.4	4.3
93792_Lupus Kidney	6.6	6.4
93577_NCI-H292	32.8	27.4

93358_NCI-H292_IL-4	41.8	34.9
93360_NCI-H292_IL-9	33.4	29.7
93359_NCI-H292_IL-13	27.0	21.9
93357_NCI-H292_IFN gamma	18.6	15.1
93777_HPAEC -	7.5	4.6
93778_HPAEC_IL-1 beta/TNA alpha	26.6	14.5
93254_Normal Human Lung Fibroblast_none	27.0	12.9
93253_Normal Human Lung Fibroblast_TNFa (4 ng/ml) and IL-1b (1 ng/ml)	11.7	7.5
93257_Normal Human Lung Fibroblast_IL-4	31.9	15.9
93256_Normal Human Lung Fibroblast_IL-9	0.0	12.8
93255_Normal Human Lung Fibroblast_IL-13	38.7	26.6
93258_Normal Human Lung Fibroblast_IFN gamma	52.5	26.1
93106_Dermal Fibroblasts CCD1070_resting	57.0	34.9
93361_Dermal Fibroblasts CCD1070_TNF alpha 4 ng/ml	87.1	60.7
93105_Dermal Fibroblasts CCD1070_IL-1 beta 1 ng/ml	46.7	37.4
93772_dermal fibroblast_IFN gamma	28.9	14.5
93771_dermal fibroblast_IL-4	51.0	27.7
93259_IBD Colitis 1**	4.9	4.8
93260_IBD Colitis 2	1.1	0.8
93261_IBD Crohns	1.0	1.0
735010_Colon_normal	12.4	6.7
735019_Lung_none	24.7	11.7
64028-1_Thymus_none	33.4	27.5
64030-1_Kidney_none	22.5	20.6

Panel 1 Summary: Ag267 Among the normal tissues on this panel, highest expression of the MOL1a gene is detected in testis (CT value = 25) and adipose. High expression in adipose might suggest that the MOL1a gene plays a role in the development of metabolic diseases, such as obesity or diabetes. In addition, expression of this gene is high in a renal cancer cell line (CT value = 25). Moderate expression of the MOL1a gene is also seen in most regions of normal brain. Strikingly, the MOL1a transcript appears at much higher levels in a number of CNS cancer cell lines. Therefore, inhibition of the MOL1a gene product using a monoclonal antibody and/or small molecule therapeutic may be useful for the treatment of some renal cell and CNS carcinomas.

Panel 1.2 Summary: Ag267 Expression of the MOL1a gene is highest in the cerebral cortex (CT value = 25) with more moderate expression detected in most other regions of normal brain, suggesting a role for this gene in neurological function. Consistent with the results seen in Panel 1, this gene is strikingly overexpressed in a number of CNS cancer cell lines (specifically glioma and astrocytoma). Moderate overexpression of the MOL1a gene is also detected in renal cell cancer and lung cancer cell lines relative to the normal controls. The

MOL1a gene product displays moderate similarity to the Notch protein that has been shown to be involved in cell signalling and has been implicated in oncogenesis. Therefore, the MOL1a gene may be a good marker for CNS or other cancers and would potentially serve as a good drug target for the treatment of certain cancers. This gene is also well expressed in several metabolic tissues (specifically adipose, liver and pancreas) and may thus have application for the treatment of metabolic diseases such as diabetes and obesity. Of particular interest is the good expression (CT value = 30.6) in pancreas. The human pancreas-specific gene SEL-11 is thought to be a negative regulator of the notch receptor (Harada, Y. et al. J Hum Genet 44(5):330-6, 1999). Thus, the notch-like MOL1a gene and notch receptor may have potential therapeutic use in diseases involving the pancreas.

Panel 1.3D Summary: Ag267 Among normal tissues, highest MOL1a transcript levels are found in adipose (CT value = 30). As was seen for Panels 1 and 1.2, moderate expression of this gene is detected in most regions of normal brain and the gene is strikingly over expressed in a number of CNS cancer cell lines. In general, expression of the MOL1a gene appears to be higher in cell lines when compared to tissue samples. A cluster of expression associated with brain, breast and renal cancer cell lines is evident. Thus, the expression of this gene could be associated with cancer cells when compared to normal, since these cell lines are derived from cancers. Alternatively, the expression of this gene could be associated with cell division, since a high percentage of cells in culture are actively dividing when compared to cells in tissue.

Panel 2D Summary: Ag267 Expression of the MOL1a gene in panel 2D appears to be widespread across most of the samples. However, there seems to be significant dysregulation in breast cancers when compared to normal adjacent tissues. Thus, therapeutic modulation of this gene might show utility in the treatment of breast cancers.

Panel 2.2 Summary: Ag1308 The expression of this gene appears to be widespread across most of the samples in panel 2.2. In a couple of instances of renal cell cancer, there seems to be significant dysregulation of the expression of this gene when compared to normal adjacent tissue. Thus, therapeutic modulation of this gene might be useful in the treatment of a sub-set of renal cancers.

Panel 4D Summary: Ag267/Ag1308 The MOL1a transcript is broadly expressed in fibroblasts, keratinocytes, B cells, and T cells, although at a moderate level. High expression of the transcript is also found in monocytes, whether activated or not. In addition, the transcript is up-regulated (7 fold) in keratinocytes and small airway epitheliun by treatment with TNFa and IL-1. The Notch-like protein encoded by the MOL1a gene may regulate cell

survival based on its homology to other Notch proteins. Therefore, protein therapeutics (agonist or antagonists) against the MOL1a gene product may be beneficial in the treatment of lung diseases, such as asthma and emphysema, or in the treatment of skin diseases, such as psoriasis and contact sensitivity.

5

B. MOL2

Expression of gene MOL2 was assessed using the primer-probe set Ag2120, described in Table 17. Results of the RTQ-PCR runs are shown in Tables 18, 19, 20, and 21

Table 17. Probe Name Ag2120

Primers	Sequences	TM	Length	Start Position	SEQ ID NO:
Forward	5'-GCTGATTGCAAGAAGATGTTTC-3'	59	22	103	71
Probe	TET-5'-TTTTGTCAGCCCTGATTTCTTCGACA-3'-TAMRA	68.8	26	140	72
Reverse	5'-CCGATATGTCAGAATCTGCATT-3'	59.1	22	166	73

Table 18. Panel 1.3D

Tissue Name	Relative Expression(%) 1.3Dtm3025t_ ag2120	Relative Expression(%) 1.3dtm3058t_ ag2120
Liver adenocarcinoma	0.5	0.1
Pancreas	0.3	0.2
Pancreatic ca. CAPAN 2	1.0	1.1
Adrenal gland	12.8	10.8
Thyroid	2.1	1.3
Salivary gland	0.6	1.1
Pituitary gland	0.6	0.3
Brain (fetal)	7.9	8.8
Brain (whole)	26.8	23.8
Brain (amygdala)	27.5	17.8
Brain (cerebellum)	17.4	19.2
Brain (hippocampus)	64.6	58.6
Brain (substantia nigra)	2.2	3.5
Brain (thalamus)	3.6	3.0
Cerebral Cortex	100.0	100.0

Spinal cord	0.5	1.2
CNS ca. (glio/astro) U87-MG	0.8	0.5
CNS ca. (glio/astro) U-118-MG	3.2	4.7
CNS ca. (astro) SW1783	0.7	0.8
CNS ca.* (neuro; met) SK-N-AS	0.4	0.2
CNS ca. (astro) SF-539	0.4	0.3
CNS ca. (astro) SNB-75	0.5	0.2
CNS ca. (glio) SNB-19	0.4	1.2
CNS ca. (glio) U251	1.6	1.3
CNS ca. (glio) SF-295	1.9	0.8
Heart (fetal)	3.3	2.2
Heart	0.3	0.6
Fetal Skeletal	6.8	9.3
Skeletal muscle	3.4	4.0
Bone marrow	0.9	1.6
Thymus	0.9	0.5
Spleen	0.4	0.0
Lymph node	1.2	0.6
Colorectal	1.6	1.1
Stomach	1.0	0.2
Small intestine	0.2	2.0
Colon ca. SW480	0.8	1.6
Colon ca.* (SW480 met)SW620	0.0	1.4
Colon ca. HT29	0.2	0.5
Colon ca. HCT-116	0.3	0.2
Colon ca. CaCo-2	0.6	0.0
83219 CC Well to Mod Diff (ODO3866)	1.5	0.5
Colon ca. HCC-2998	0.6	0.3
Gastric ca.* (liver met) NCI-N87	0.5	1.2
Bladder	1.8	1.0
Trachea	0.6	1.3
Kidney	1.3	2.5
Kidney (fetal)	1.7	4.0
Renal ca. 786-0	3.8	3.7
Renal ca. A498	2.6	3.0
Renal ca. RXF 393	2.2	2.3
Renal ca. ACHN	1.1	0.8
Renal ca. UO-31	0.5	1.2
Renal ca. TK-10	3.7	3.6
Liver	0.8	0.0
Liver (fetal)	0.2	0.0
Liver ca. (hepatoblast) HepG2	0.7	0.8
Lung	0.2	0.7
Lung (fetal)	0.4	0.6
Lung ca. (small cell) LX-1	1.2	0.1

Lung ca. (small cell) NCI-H69	44.8	40.9
Lung ca. (s.cell var.) SHP-77	5.6	5.2
Lung ca. (large cell)NCI-H460	54.3	74.2
Lung ca. (non-sm. cell) A549	0.5	0.0
Lung ca. (non-s.cell) NCI-H23	1.5	2.0
Lung ca (non-s.cell) HOP-62	7.0	5.3
Lung ca. (non-s.cl) NCI-H522	1.1	0.0
Lung ca. (squam.) SW 900	0.3	0.8
Lung ca. (squam.) NCI-H596	8.4	11.1
Mammary gland	2.4	2.0
Breast ca.* (pl. effusion) MCF-7	1.0	0.7
Breast ca.* (pl.ef) MDA-MB-231	0.0	0.0
Breast ca.* (pl. effusion) T47D	17.3	18.7
Breast ca. BT-549	4.2	4.7
Breast ca. MDA-N	9.3	9.9
Ovary	12.4	9.9
Ovarian ca. OVCAR-3	0.1	0.0
Ovarian ca. OVCAR-4	0.5	2.7
Ovarian ca. OVCAR-5	0.9	0.9
Ovarian ca. OVCAR-8	2.1	4.0
Ovarian ca. IGROV-1	0.4	0.3
Ovarian ca.* (ascites) SK-OV-3	0.8	0.0
Uterus	1.2	0.7
Placenta	1.8	0.1
Prostate	0.5	0.6
Prostate ca.* (bone met)PC-3	3.1	2.2
Testis	1.2	1.3
Melanoma Hs688(A).T	0.8	0.9
Melanoma* (met) Hs688(B).T	1.2	2.7
Melanoma UACC-62	0.7	0.7
Melanoma M14	4.0	5.5
Melanoma LOX IMVI	1.6	0.3
Melanoma* (met) SK-MEL-5	4.2	3.7
Adipose	3.1	1.9

Table 19. Panel 2D

Tissue Name	Relative Expression(%)	
	2Dtm3026t_ ag2120	2Dtm3035t_ ag2120
Normal Colon GENPAK 061003	7.3	7.4
83219 CC Well to Mod Diff (ODO3866)	4.3	7.9
83220 CC NAT (ODO3866)	1.7	2.9

83221 CC Gr.2 rectosigmoid (ODO3868)	2.3	1.5
83222 CC NAT (ODO3868)	1.7	3.8
83235 CC Mod Diff (ODO3920)	0.2	0.9
83236 CC NAT (ODO3920)	0.1	2.5
83237 CC Gr.2 ascend colon (ODO3921)	0.5	0.4
83238 CC NAT (ODO3921)	5.8	6.4
83241 CC from Partial Hepatectomy (ODO4309)	6.1	8.7
83242 Liver NAT (ODO4309)	4.0	2.2
87472 Colon mets to lung (OD04451-01)	0.0	0.3
87473 Lung NAT (OD04451-02)	2.9	3.9
Normal Prostate Clontech A+ 6546-1	3.0	7.8
84140 Prostate Cancer (OD04410)	8.5	9.0
84141 Prostate NAT (OD04410)	23.5	21.5
87073 Prostate Cancer (OD04720-01)	7.3	7.1
87074 Prostate NAT (OD04720-02)	1.7	11.3
Normal Lung GENPAK 061010	9.2	7.2
83239 Lung Met to Muscle (ODO4286)	0.7	0.7
83240 Muscle NAT (ODO4286)	11.3	12.5
84136 Lung Malignant Cancer (OD03126)	7.9	4.7
84137 Lung NAT (OD03126)	6.4	7.5
84871 Lung Cancer (OD04404)	3.0	3.8
84872 Lung NAT (OD04404)	2.6	2.8
84875 Lung Cancer (OD04565)	1.8	3.4
84876 Lung NAT (OD04565)	3.6	2.8
85950 Lung Cancer (OD04237-01)	22.5	17.3
85970 Lung NAT (OD04237-02)	3.5	4.5
83255 Ocular Mel Met to Liver (ODO4310)	2.1	4.1
83256 Liver NAT (ODO4310)	1.2	1.2
84139 Melanoma Mets to Lung (OD04321)	2.5	1.7
84138 Lung NAT (OD04321)	5.3	3.3
Normal Kidney GENPAK 061008	93.3	**100.0**
83786 Kidney Ca, Nuclear grade 2 (OD04338)	55.9	92.7
83787 Kidney NAT (OD04338)	37.9	36.6
83788 Kidney Ca Nuclear grade 1/2 (OD04339)	67.4	76.3
83789 Kidney NAT (OD04339)	25.3	33.2
83790 Kidney Ca, Clear cell type (OD04340)	52.5	43.5
83791 Kidney NAT (OD04340)	0.0	35.6
83792 Kidney Ca, Nuclear grade 3 (OD04348)	5.0	7.6
83793 Kidney NAT (OD04348)	20.7	26.6
87474 Kidney Cancer (OD04622-01)	9.3	7.3
87475 Kidney NAT (OD04622-03)	7.1	10.2
85973 Kidney Cancer (OD04450-01)	27.0	29.3
85974 Kidney NAT (OD04450-03)	34.2	33.7
Kidney Cancer Clontech 8120607	4.1	3.3
Kidney NAT Clontech 8120608	9.2	12.8

Kidney Cancer Clontech 8120613	2.2	3.3
Kidney NAT Clontech 8120614	7.3	14.9
Kidney Cancer Clontech 9010320	22.4	26.4
Kidney NAT Clontech 9010321	18.3	26.8
Normal Uterus GENPAK 061018	9.6	8.6
Uterus Cancer GENPAK 064011	2.7	2.6
Normal Thyroid Clontech A+ 6570-1	5.1	5.3
Thyroid Cancer GENPAK 064010	39.2	44.1
Thyroid Cancer INVITROGEN A302152	30.8	24.5
Thyroid NAT INVITROGEN A302153	3.3	3.6
Normal Breast GENPAK 061019	5.0	4.5
84877 Breast Cancer (OD04566)	0.8	1.7
85975 Breast Cancer (OD04590-01)	9.7	7.6
85976 Breast Cancer Mets (OD04590-03)	26.1	33.4
87070 Breast Cancer Metastasis (OD04655-05)	3.4	4.1
GENPAK Breast Cancer 064006	3.3	4.4
Breast Cancer Res. Gen. 1024	7.2	8.3
Breast Cancer Clontech 9100266	3.1	3.7
Breast NAT Clontech 9100265	3.3	3.9
Breast Cancer INVITROGEN A209073	8.2	8.9
Breast NAT INVITROGEN A2090734	16.2	11.6
Normal Liver GENPAK 061009	1.4	0.9
Liver Cancer GENPAK 064003	1.8	5.1
Liver Cancer Research Genetics RNA 1025	1.3	2.0
Liver Cancer Research Genetics RNA 1026	3.4	2.4
Paired Liver Cancer Tissue Research Genetics RNA 6004-T	1.6	0.4
Paired Liver Tissue Research Genetics RNA 6004-N	1.3	1.0
Paired Liver Cancer Tissue Research Genetics RNA 6005-T	0.0	6.5
Paired Liver Tissue Research Genetics RNA 6005-N	1.6	0.3
Normal Bladder GENPAK 061001	3.4	9.5
Bladder Cancer Research Genetics RNA 1023	10.7	6.5
Bladder Cancer INVITROGEN A302173	1.8	2.1
87071 Bladder Cancer (OD04718-01)	1.7	5.2
87072 Bladder Normal Adjacent (OD04718-03)	4.2	6.3
Normal Ovary Res. Gen.	9.5	7.4
Ovarian Cancer GENPAK 064008	100.0	95.3
87492 Ovary Cancer (OD04768-07)	4.6	7.6
87493 Ovary NAT (OD04768-08)	7.3	5.7
Normal Stomach GENPAK 061017	1.7	3.8
Gastric Cancer Clontech 9060358	2.1	0.5
NAT Stomach Clontech 9060359	1.6	3.1
Gastric Cancer Clontech 9060395	3.3	3.1
NAT Stomach Clontech 9060394	2.2	3.3
Gastric Cancer Clontech 9060397	11.0	14.5
NAT Stomach Clontech 9060396	2.8	4.3

Gastric Cancer GENPAK 064005	2.2	6.5

Table 20. Panel 4D

Tissue Name	Relative Expression(%) 4Dtm3027t_ag2120	Tissue Name	Relative Expression(%) 4Dtm3027t_ag2120
93768_Secondary Th1_anti-CD28/anti-CD3	0.4	93100_HUVEC (Endothelial)_IL-1b	11.2
93769_Secondary Th2_anti-CD28/anti-CD3	0.8	93779_HUVEC (Endothelial)_IFN gamma	10.1
93770_Secondary Tr1_anti-CD28/anti-CD3	3.1	93102_HUVEC (Endothelial)_TNF alpha + IFN gamma	4.1
93573_Secondary Th1_resting day 4-6 in IL-2	3.4	93101_HUVEC (Endothelial)_TNF alpha + IL4	13.2
93572_Secondary Th2_resting day 4-6 in IL-2	1.5	93781_HUVEC (Endothelial)_IL-11	8.1
93571_Secondary Tr1_resting day 4-6 in IL-2	1.5	93583_Lung Microvascular Endothelial Cells_none	3.2
93568_primary Th1_anti-CD28/anti-CD3	0.3	93584_Lung Microvascular Endothelial Cells_TNFa (4 ng/ml) and IL1b (1 ng/ml)	2.4
93569_primary Th2_anti-CD28/anti-CD3	0.8	92662_Microvascular Dermal endothelium_none	0.0
93570_primary Tr1_anti-CD28/anti-CD3	0.4	92663_Microsvasular Dermal endothelium_TNFa (4 ng/ml) and IL1b (1 ng/ml)	0.0
93565_primary Th1_resting dy 4-6 in IL-2	3.1	93773_Bronchial epithelium_TNFa (4 ng/ml) and IL1b (1 ng/ml) **	3.1
93566_primary Th2_resting dy 4-6 in IL-2	5.5	93347_Small Airway Epithelium_none	5.4
93567_primary Tr1_resting dy 4-6 in IL-2	0.6	93348_Small Airway Epithelium_TNFa (4 ng/ml) and IL1b (1 ng/ml)	3.0
93351_CD45RA CD4 lymphocyte_anti-CD28/anti-CD3	1.3	92668_Coronery Artery SMC_resting	35.6
93352_CD45RO CD4 lymphocyte_anti-CD28/anti-CD3	2.4	92669_Coronery Artery SMC_TNFa (4 ng/ml) and IL1b (1 ng/ml)	39.0
93251_CD8 Lymphocytes_anti-CD28/anti-CD3	2.2	93107_astrocytes_resting	1.1
93353_chronic CD8 Lymphocytes 2ry_resting dy 4-6 in IL-2	4.8	93108_astrocytes_TNFa (4 ng/ml) and IL1b (1 ng/ml)	0.9
93574_chronic CD8 Lymphocytes 2ry_activated	2.4	92666_KU-812 (Basophil)_resting	0.5

CD3/CD28			
93354_CD4_none	0.9	92667_KU-812 (Basophil)_PMA/ionoycin	2.0
93252_Secondary Th1/Th2/Tr1_anti-CD95 CH11	0.7	93579_CCD1106 (Keratinocytes)_none	0.9
93103_LAK cells_resting	7.2	93580_CCD1106 (Keratinocytes)_TNFa and IFNg **	2.6
93788_LAK cells_IL-2	0.4	93791_Liver Cirrhosis	5.5
93787_LAK cells_IL-2+IL-12	2.8	93792_Lupus Kidney	9.3
93789_LAK cells_IL-2+IFN gamma	7.7	93577_NCI-H292	2.0
93790_LAK cells_IL-2+ IL-18	4.1	93358_NCI-H292_IL-4	5.3
93104_LAK cells_PMA/ionomycin and IL-18	3.2	93360_NCI-H292_IL-9	3.8
93578_NK Cells IL-2_resting	2.3	93359_NCI-H292_IL-13	1.3
93109_Mixed Lymphocyte Reaction_Two Way MLR	2.7	93357_NCI-H292_IFN gamma	3.4
93110_Mixed Lymphocyte Reaction_Two Way MLR	4.0	93777_HPAEC -	9.0
93111_Mixed Lymphocyte Reaction_Two Way MLR	5.2	93778_HPAEC_IL-1 beta/TNA alpha	15.4
93112_Mononuclear Cells (PBMCs)_resting	6.4	93254_Normal Human Lung Fibroblast_none	1.7
93113_Mononuclear Cells (PBMCs)_PWM	5.2	93253_Normal Human Lung Fibroblast_TNFa (4 ng/ml) and IL-1b (1 ng/ml)	0.0
93114_Mononuclear Cells (PBMCs)_PHA-L	1.8	93257_Normal Human Lung Fibroblast_IL-4	2.8
93249_Ramos (B cell)_none	3.9	93256_Normal Human Lung Fibroblast_IL-9	1.6
93250_Ramos (B cell)_ionomycin	1.6	93255_Normal Human Lung Fibroblast_IL-13	1.0
93349_B lymphocytes_PWM	0.0	93258_Normal Human Lung Fibroblast_IFN gamma	1.4
93350_B lymphoytes_CD40L and IL-4	3.2	93106_Dermal Fibroblasts CCD1070_resting	0.6
92665_EOL-1 (Eosinophil)_dbcAMP differentiated	67.4	93361_Dermal Fibroblasts CCD1070_TNF alpha 4 ng/ml	2.3
93248_EOL-1 (Eosinophil)_dbcAMP/PMAionomycin	56.6	93105_Dermal Fibroblasts CCD1070_IL-1 beta 1 ng/ml	5.4
93356_Dendritic Cells_none	6.0	93772_dermal fibroblast_IFN gamma	0.5
93355_Dendritic Cells_LPS 100 ng/ml	15.5	93771_dermal fibroblast_IL-4	3.0
93775_Dendritic Cells_anti-CD40	13.5	93259_IBD Colitis 1**	1.3
93774_Monocytes_resting	46.0	93260_IBD Colitis 2	5.8
93776_Monocytes_LPS 50	4.9	93261_IBD Crohns	4.2

ng/ml			
93581_Macrophages_resting	47.3	735010_Colon_normal	5.8
93582_Macrophages_LPS 100 ng/ml	24.8	735019_Lung_none	7.0
93098_HUVEC (Endothelial)_none	10.2	64028-1_Thymus_none	**100.0**
93099_HUVEC (Endothelial)_starved	23.8	64030-1_Kidney_none	18.2

Table 21. Panel CNSD.01

Tissue Name	Relative Expression(%) CNS1x4tm6184t_ag2120_a2	Tissue Name	Relative Expression(%) CNS1x4tm6184t_ag2120_a2
102633_BA4 Control	46.5	102605_BA17 PSP	20.1
102641_BA4 Control2	42.9	102612_BA17 PSP2	10.5
102625_BA4 Alzheimer's2	5.2	102637_Sub Nigra Control	10.6
102649_BA4 Parkinson's	38.5	102645_Sub Nigra Control2	29.1
102656_BA4 Parkinson's2	**100.0**	102629_Sub Nigra Alzheimer's2	5.8
102664_BA4 Huntington's	53.8	102660_Sub Nigra Parkinson's2	20.6
102671_BA4 Huntington's2	4.3	102667_Sub Nigra Huntington's	21.5
102603_BA4 PSP	4.5	102674_Sub Nigra Huntington's2	14.7
102610_BA4 PSP2	14.1	102614_Sub Nigra PSP2	3.6
102588_BA4 Depression	11.8	102592_Sub Nigra Depression	3.2
102596_BA4 Depression2	3.5	102599_Sub Nigra Depression2	2.4
102634_BA7 Control	45.6	102636_Glob Palladus Control	1.8
102642_BA7 Control2	37.9	102644_Glob Palladus Control2	7.8
102626_BA7 Alzheimer's2	8.5	102620_Glob Palladus Alzheimer's	6.4
102650_BA7 Parkinson's	11.7	102628_Glob Palladus Alzheimer's2	2.5
102657_BA7 Parkinson's2	54.6	102652_Glob Palladus Parkinson's	34.4
102665_BA7 Huntington's	41.1	102659_Glob Palladus Parkinson's2	3.1
102672_BA7 Huntington's2	34.3	102606_Glob Palladus PSP	5.2
102604_BA7 PSP	36.2	102613_Glob Palladus PSP2	0.0
102611_BA7 PSP2	21.2	102591_Glob Palladus Depression	3.7
102589_BA7 Depression	4.5	102638_Temp Pole Control	11.0
102632_BA9 Control	18.1	102646_Temp Pole Control2	40.1
102640_BA9 Control2	76.7	102622_Temp Pole Alzheimer's	2.7
102617_BA9 Alzheimer's	8.0	102630_Temp Pole	3.4

		Alzheimer's2	
102624_BA9 Alzheimer's2	4.6	102653_Temp Pole Parkinson's	13.9
102648_BA9 Parkinson's	28.7	102661_Temp Pole Parkinson's2	20.8
102655_BA9 Parkinson's2	55.5	102668_Temp Pole Huntington's	27.6
102663_BA9 Huntington's	34.2	102607_Temp Pole PSP	2.5
102670_BA9 Huntington's2	4.7	102615_Temp Pole PSP2	3.7
102602_BA9 PSP	11.6	102600_Temp Pole Depression2	1.8
102609_BA9 PSP2	3.8	102639_Cing Gyr Control	70.7
102587_BA9 Depression	5.0	102647_Cing Gyr Control2	23.9
102595_BA9 Depression2	3.2	102623_Cing Gyr Alzheimer's	13.8
102635_BA17 Control	37.0	102631_Cing Gyr Alzheimer's2	7.8
102643_BA17 Control2	62.0	102654_Cing Gyr Parkinson's	13.2
102627_BA17 Alzheimer's2	2.9	102662_Cing Gyr Parkinson's2	23.0
102651_BA17 Parkinson's	13.4	102669_Cing Gyr Huntington's	40.3
102658_BA17 Parkinson's2	61.3	102676_Cing Gyr Huntington's2	12.9
102666_BA17 Huntington's	28.0	102608_Cing Gyr PSP	7.6
102673_BA17 Huntington's2	8.4	102616_Cing Gyr PSP2	5.6
102590_BA17 Depression	2.3	102594_Cing Gyr Depression	4.1
102597_BA17 Depression2	10.1	102601_Cing Gyr Depression2	4.6

Panel 1.3D Summary: Ag2120 Two replicate experiments using the same probe and primer set show very comparable results. Expression of the MOL2 gene is highest in the cerebral cortex (CT value = 29). Moderate expression is detected in all other regions of the brain except thalamus and substantia nigra; this observation suggests that the MOL2 gene may be associated with normal brain homeostasis. Thus, this protein shows a brain-preferential expression; see write-up on Panel CNS.01 for discussion of utility. In addition, expression of the MOL2 gene appears to be down-regulated in CNS cancer cell lines. Overexpression of the MOL2 gene is also detected in several lung cancer cell lines relative to normal control. Therefore, this gene might be a good target for the detection or treatment of CNS and lung cancers.

Panel 2D Summary: Ag2120 Two replicate experiments using the same probe and primer set show very comparable results. Expression of the MOL2 gene in panel 2D reveals an association of expression in thyroid, breast and kidney cancers when compared to their respective normal adjacent tissues. Thus, therapeutic modulation of this gene with inhibitory monoclonal antibodies and/or small moleculte therapeutics may show utility in treatment of

these diseases. In addition, the MOL2 gene might be useful as a marker for thyroid, breast and kidney cancers.

Panel 4D Summary: Ag 2120 The MOL2 gene is expressed at highest levels in the thymus (CT value =31), In addition, the transcript is also expressed in eosinophils, monocytes, macrophages and coronary artery. Interestingly, it is down regulated in LPS-treated monocytes and to a lesser degree in LPS treated macrophages. Therefore, protein therapeutics (agonists or antagonists) designed against the protein encoded for by this transcript could reduce inflammatory process observed in asthma, emphysema, osteoarthritis and sepsis.

Panel CNSD.01 Summary: Ag2120 The insulin and insulin-like growth factors belong to a family of polypeptides essential for proper regulation of physiologic processes such as energy metabolism, cell proliferation, development, and differentiation. The insulin-like growth factors bind to IGF with high affinity and compete with the IGF receptor for IGF binding. Transgenic mice overexpressing insulin-like growth factor binding proteins (IGFBPs) tend to show brain developmental abnormalities, suggesting a role for these proteins in neurodevelopment. Furthermore, treatment with glycosaminoglycans (which increases muscle re-innervation after motor neuron death) upregulates serum levels of both IGF and IGFBP. Thus, the novel IGFBP encoded by the MOL2 gene may be useful in the treatment of diseases such as ALS, multiple sclerosis, and peripheral nerve injury on the basis of its homology to other established IGFBPs. The expression profile of this gene suggests that it is expressed preferentially in the brain, with highest levels in the cerebral cortex and hippocampus, two regions that are known to degenerate in Alzheimer's disease. Examination of the expression profile on Panel CNS.01 shows that most regions of both control and diseased brains express this protein; however the levels are decreased in the motor cortex in progressive supranuclear palsy and depression. Thus, this protein may additionally be of use in the treatment of Alzheimer's disease, progressive supranuclear palsy, and depression.

C. MOL3a

Expression of gene MOL3a was assessed using the primer-probe set Ag1493, described in Table 22. Results of the RTQ-PCR runs are shown in Tables 23, 24, 25, and 26.

Table 22. Probe Name Ag1493

Primers	Sequences	TM	Length	Start Position	SEQ ID NO:
Forward	5'-GTGAAATCTGGCGTGGAGTATA-3'	59.1	22	1224	74
Probe	FAM-5'-CCTTGATGGGCACAGCCATCTTG-3'-TAMRA	70	23	1274	75
Reverse	5'-GTACTGGTTCCCAGGTACATGA-3'	58.8	22	1318	76

Table 23. Panel 1.2

Tissue Name	Relative Expression(%) 1.2tm2058f_ag1493	Tissue Name	Relative Expression(%) 1.2tm2058f_ag1493
Endothelial cells	0.2	Renal ca. 786-0	0.0
Endothelial cells (treated)	2.4	Renal ca. A498	0.5
Pancreas	0.1	Renal ca. RXF 393	0.9
Pancreatic ca. CAPAN 2	0.1	Renal ca. ACHN	0.2
Adrenal Gland (new lot*)	2.6	Renal ca. UO-31	1.3
Thyroid	1.0	Renal ca. TK-10	0.8
Salivary gland	21.8	Liver	1.4
Pituitary gland	0.4	Liver (fetal)	2.3
Brain (fetal)	3.3	Liver ca. (hepatoblast) HepG2	0.7
Brain (whole)	2.1	Lung	0.8
Brain (amygdala)	8.0	Lung (fetal)	1.0
Brain (cerebellum)	0.3	Lung ca. (small cell) LX-1	0.4
Brain (hippocampus)	15.0	Lung ca. (small cell) NCI-H69	1.9
Brain (thalamus)	3.5	Lung ca. (s.cell var.) SHP-77	0.0
Cerebral Cortex	30.1	Lung ca. (large cell)NCI-H460	0.7
Spinal cord	0.2	Lung ca. (non-sm. cell) A549	1.0
CNS ca. (glio/astro) U87-MG	0.2	Lung ca. (non-s.cell) NCI-H23	0.2
CNS ca. (glio/astro) U-118-MG	0.3	Lung ca (non-s.cell) HOP-62	1.8
CNS ca. (astro) SW1783	0.0	Lung ca. (non-s.cl) NCI-H522	11.9
CNS ca.* (neuro; met) SK-N-AS	0.5	Lung ca. (squam.) SW 900	0.8
CNS ca. (astro) SF-539	0.4	Lung ca. (squam.) NCI-H596	1.0
CNS ca. (astro) SNB-75	0.1	Mammary gland	2.5
CNS ca. (glio) SNB-19	0.4	Breast ca.* (pl. effusion) MCF-7	3.7
CNS ca. (glio) U251	0.2	Breast ca.* (pl.ef) MDA-MB-231	0.0
CNS ca. (glio) SF-295	1.3	Breast ca.* (pl. effusion) T47D	17.6
Heart	5.1	Breast ca. BT-549	0.2
Skeletal Muscle (new lot*)	8.0	Breast ca. MDA-N	0.4
Bone marrow	15.6	Ovary	2.7
Thymus	1.2	Ovarian ca. OVCAR-3	8.3
Spleen	10.4	Ovarian ca. OVCAR-4	16.6

Lymph node	0.7	Ovarian ca. OVCAR-5	5.1
Colorectal	0.9	Ovarian ca. OVCAR-8	0.8
Stomach	2.9	Ovarian ca. IGROV-1	70.7
Small intestine	4.7	Ovarian ca.* (ascites) SK-OV-3	0.7
Colon ca. SW480	0.2	Uterus	0.6
Colon ca.* (SW480 met)SW620	0.0	Placenta	0.6
Colon ca. HT29	2.3	Prostate	27.9
Colon ca. HCT-116	0.0	Prostate ca.* (bone met)PC-3	0.3
Colon ca. CaCo-2	0.0	Testis	0.6
83219 CC Well to Mod Diff (ODO3866)	1.4	Melanoma Hs688(A).T	0.2
Colon ca. HCC-2998	2.8	Melanoma* (met) Hs688(B).T	0.4
Gastric ca.* (liver met) NCI-N87	1.6	Melanoma UACC-62	0.5
Bladder	6.9	Melanoma M14	0.1
Trachea	0.5	Melanoma LOX IMVI	0.1
Kidney	73.7	Melanoma* (met) SK-MEL-5	0.0
Kidney (fetal)	4.7	Adipose	**100.0**

Table 24. Panel 1.3D

Tissue Name	Relative Expression(%) 1.3dx4tm5350f ag1493_b1	Tissue Name	Relative Expression(%) 1.3dx4tm5350f ag1493_b1
Liver adenocarcinoma	1.2	Kidney (fetal)	18.7
Pancreas	0.0	Renal ca. 786-0	0.0
Pancreatic ca. CAPAN 2	1.9	Renal ca. A498	0.0
Adrenal gland	1.7	Renal ca. RXF 393	2.9
Thyroid	38.0	Renal ca. ACHN	1.0
Salivary gland	30.8	Renal ca. UO-31	0.8
Pituitary gland	4.5	Renal ca. TK-10	0.8
Brain (fetal)	48.3	Liver	0.4
Brain (whole)	60.7	Liver (fetal)	11.2
Brain (amygdala)	**100.0**	Liver ca. (hepatoblast) HepG2	0.6
Brain (cerebellum)	9.2	Lung	29.6
Brain (hippocampus)	68.2	Lung (fetal)	15.6
Brain (substantia nigra)	11.4	Lung ca. (small cell) LX-1	0.2
Brain (thalamus)	20.8	Lung ca. (small cell) NCI-H69	0.0
Cerebral Cortex	41.1	Lung ca. (s.cell var.) SHP-77	0.0
Spinal cord	10.0	Lung ca. (large cell)NCI-H460	0.5
CNS ca. (glio/astro) U87-MG	0.0	Lung ca. (non-sm. cell) A549	0.4
CNS ca. (glio/astro) U-118-MG	1.3	Lung ca. (non-s.cell) NCI-H23	0.0
CNS ca. (astro) SW1783	0.7	Lung ca (non-s.cell) HOP-62	2.6
CNS ca.* (neuro; met) SK-N-AS	1.9	Lung ca. (non-s.cl) NCI-H522	3.0

CNS ca. (astro) SF-539	0.5	Lung ca. (squam.) SW 900	0.2
CNS ca. (astro) SNB-75	0.4	Lung ca. (squam.) NCI-H596	0.0
CNS ca. (glio) SNB-19	0.4	Mammary gland	22.3
CNS ca. (glio) U251	1.2	Breast ca.* (pl. effusion) MCF-7	14.1
CNS ca. (glio) SF-295	0.6	Breast ca.* (pl.ef) MDA-MB-231	0.0
Heart (fetal)	1.3	Breast ca.* (pl. effusion) T47D	53.3
Heart	1.4	Breast ca. BT-549	0.5
Fetal Skeletal	4.2	Breast ca. MDA-N	0.0
Skeletal muscle	8.8	Ovary	2.1
Bone marrow	78.4	Ovarian ca. OVCAR-3	8.7
Thymus	3.9	Ovarian ca. OVCAR-4	25.1
Spleen	53.3	Ovarian ca. OVCAR-5	3.8
Lymph node	37.3	Ovarian ca. OVCAR-8	2.6
Colorectal	6.6	Ovarian ca. IGROV-1	29.2
Stomach	23.1	Ovarian ca.* (ascites) SK-OV-3	0.5
Small intestine	12.5	Uterus	8.9
Colon ca. SW480	0.0	Placenta	10.9
Colon ca.* (SW480 met)SW620	0.0	Prostate	50.9
Colon ca. HT29	2.9	Prostate ca.* (bone met)PC-3	0.0
Colon ca. HCT-116	0.0	Testis	18.7
Colon ca. CaCo-2	0.8	Melanoma Hs688(A).T	0.5
83219 CC Well to Mod Diff (ODO3866)	8.2	Melanoma* (met) Hs688(B).T	0.0
Colon ca. HCC-2998	0.8	Melanoma UACC-62	0.0
Gastric ca.* (liver met) NCI-N87	2.8	Melanoma M14	0.5
Bladder	5.3	Melanoma LOX IMVI	0.0
Trachea	35.6	Melanoma* (met) SK-MEL-5	0.0
Kidney	15.0	Adipose	60.2

Table 25. Panel 2D

Tissue Name	Relative Expression(%) 2Dtm2527f_ag 1493	Tissue Name	Relative Expression(%) 2Dtm2527f_ag 1493
Normal Colon GENPAK 061003	41.2	Kidney NAT Clontech 8120608	28.3
83219 CC Well to Mod Diff (ODO3866)	8.0	Kidney Cancer Clontech 8120613	10.7
83220 CC NAT (ODO3866)	9.3	Kidney NAT Clontech 8120614	46.0
83221 CC Gr.2 rectosigmoid (ODO3868)	5.1	Kidney Cancer Clontech 9010320	48.6
83222 CC NAT (ODO3868)	2.0	Kidney NAT Clontech 9010321	36.6
83235 CC Mod Diff	6.3	Normal Uterus GENPAK	8.4

(ODO3920)		061018	
83236 CC NAT (ODO3920)	4.4	Uterus Cancer GENPAK 064011	17.4
83237 CC Gr.2 ascend colon (ODO3921)	12.5	Normal Thyroid Clontech A+ 6570-1	**100.0**
83238 CC NAT (ODO3921)	6.1	Thyroid Cancer GENPAK 064010	39.8
83241 CC from Partial Hepatectomy (ODO4309)	17.0	Thyroid Cancer INVITROGEN A302152	28.3
83242 Liver NAT (ODO4309)	7.6	Thyroid NAT INVITROGEN A302153	44.4
87472 Colon mets to lung (OD04451-01)	16.6	Normal Breast GENPAK 061019	36.3
87473 Lung NAT (OD04451-02)	25.5	84877 Breast Cancer (OD04566)	24.7
Normal Prostate Clontech A+ 6546-1	82.4	85975 Breast Cancer (OD04590-01)	29.7
84140 Prostate Cancer (OD04410)	36.3	85976 Breast Cancer Mets (OD04590-03)	17.8
84141 Prostate NAT (OD04410)	47.0	87070 Breast Cancer Metastasis (OD04655-05)	79.6
87073 Prostate Cancer (OD04720-01)	36.1	GENPAK Breast Cancer 064006	25.9
87074 Prostate NAT (OD04720-02)	51.4	Breast Cancer Res. Gen. 1024	55.5
Normal Lung GENPAK 061010	41.8	Breast Cancer Clontech 9100266	24.0
83239 Lung Met to Muscle (ODO4286)	14.4	Breast NAT Clontech 9100265	13.7
83240 Muscle NAT (ODO4286)	4.5	Breast Cancer INVITROGEN A209073	29.1
84136 Lung Malignant Cancer (OD03126)	36.1	Breast NAT INVITROGEN A2090734	29.9
84137 Lung NAT (OD03126)	71.2	Normal Liver GENPAK 061009	1.0
84871 Lung Cancer (OD04404)	68.3	Liver Cancer GENPAK 064003	1.6
84872 Lung NAT (OD04404)	33.7	Liver Cancer Research Genetics RNA 1025	7.4
84875 Lung Cancer (OD04565)	25.5	Liver Cancer Research Genetics RNA 1026	4.3
84876 Lung NAT (OD04565)	18.4	Paired Liver Cancer Tissue Research Genetics RNA 6004-T	8.5
85950 Lung Cancer (OD04237-01)	45.1	Paired Liver Tissue Research Genetics RNA 6004-N	10.1
85970 Lung NAT (OD04237-02)	55.1	Paired Liver Cancer Tissue Research Genetics RNA 6005-T	7.5
83255 Ocular Mel Met to Liver (ODO4310)	8.7	Paired Liver Tissue Research Genetics RNA 6005-N	2.8
83256 Liver NAT (ODO4310)	4.7	Normal Bladder GENPAK 061001	11.7
84139 Melanoma Mets to Lung	4.0	Bladder Cancer Research	6.4

(OD04321)		Genetics RNA 1023	
84138 Lung NAT (OD04321)	50.3	Bladder Cancer INVITROGEN A302173	50.3
Normal Kidney GENPAK 061008	53.2	87071 Bladder Cancer (OD04718-01)	35.4
83786 Kidney Ca, Nuclear grade 2 (OD04338)	13.6	87072 Bladder Normal Adjacent (OD04718-03)	13.6
83787 Kidney NAT (OD04338)	54.0	Normal Ovary Res. Gen.	6.4
83788 Kidney Ca Nuclear grade 1/2 (OD04339)	23.0	Ovarian Cancer GENPAK 064008	22.1
83789 Kidney NAT (OD04339)	26.6	87492 Ovary Cancer (OD04768-07)	29.3
83790 Kidney Ca, Clear cell type (OD04340)	31.0	87493 Ovary NAT (OD04768-08)	20.7
83791 Kidney NAT (OD04340)	49.7	Normal Stomach GENPAK 061017	21.9
83792 Kidney Ca, Nuclear grade 3 (OD04348)	7.8	Gastric Cancer Clontech 9060358	12.6
83793 Kidney NAT (OD04348)	35.8	NAT Stomach Clontech 9060359	33.7
87474 Kidney Cancer (OD04622-01)	17.8	Gastric Cancer Clontech 9060395	15.6
87475 Kidney NAT (OD04622-03)	15.4	NAT Stomach Clontech 9060394	51.8
85973 Kidney Cancer (OD04450-01)	1.7	Gastric Cancer Clontech 9060397	16.8
85974 Kidney NAT (OD04450-03)	43.2	NAT Stomach Clontech 9060396	12.4
Kidney Cancer Clontech 8120607	4.3	Gastric Cancer GENPAK 064005	12.9

Table 26. Panel 4.1D

Tissue Name	Relative Expression(%) 4.1dx4tm6520fag1493_a1	Tissue Name	Relative Expression(%) 4.1dx4tm6520fag1493_a1
93768_Secondary Th1_anti-CD28/anti-CD3	3.8	93100_HUVEC (Endothelial)_IL-1b	0.0
93769_Secondary Th2_anti-CD28/anti-CD3	27.0	93779_HUVEC (Endothelial)_IFN gamma	0.6
93770_Secondary Tr1_anti-CD28/anti-CD3	11.9	93102_HUVEC (Endothelial)_TNF alpha + IFN gamma	1.0
93573_Secondary Th1_resting day 4-6 in IL-2	1.1	93101_HUVEC (Endothelial)_TNF alpha + IL4	0.9
93572_Secondary Th2_resting day 4-6 in IL-2	3.2	93781_HUVEC (Endothelial)_IL-11	0.6
93571_Secondary Tr1_resting day 4-6 in IL-2	1.6	93583_Lung Microvascular Endothelial Cells_none	1.6
93568_primary Th1_anti-	6.7	93584_Lung Microvascular	0.6

CD28/anti-CD3		Endothelial Cells_TNFa (4 ng/ml) and IL1b (1 ng/ml)	
93569_primary Th2_anti-CD28/anti-CD3	8.4	92662_Microvascular Dermal endothelium_none	0.2
93570_primary Tr1_anti-CD28/anti-CD3	1.9	92663_Microsvasular Dermal endothelium_TNFa (4 ng/ml) and IL1b (1 ng/ml)	0.3
93565_primary Th1_resting dy 4-6 in IL-2	0.4	93773_Bronchial epithelium_TNFa (4 ng/ml) and IL1b (1 ng/ml) **	2.1
93566_primary Th2_resting dy 4-6 in IL-2	0.7	93347_Small Airway Epithelium_none	1.0
93567_primary Tr1_resting dy 4-6 in IL-2	0.2	93348_Small Airway Epithelium_TNFa (4 ng/ml) and IL1b (1 ng/ml)	4.2
93351_CD45RA CD4 lymphocyte_anti-CD28/anti-CD3	3.8	92668_Coronery Artery SMC_resting	0.1
93352_CD45RO CD4 lymphocyte_anti-CD28/anti-CD3	7.1	92669_Coronery Artery SMC_TNFa (4 ng/ml) and IL1b (1 ng/ml)	0.4
93251_CD8 Lymphocytes_anti-CD28/anti-CD3	6.1	93107_astrocytes_resting	0.1
93353_chronic CD8 Lymphocytes 2ry_resting dy 4-6 in IL-2	25.6	93108_astrocytes_TNFa (4 ng/ml) and IL1b (1 ng/ml)	0.0
93574_chronic CD8 Lymphocytes 2ry_activated CD3/CD28	9.9	92666_KU-812 (Basophil)_resting	0.3
93354_CD4_none	0.5	92667_KU-812 (Basophil)_PMA/ionoycin	0.5
93252_Secondary Th1/Th2/Tr1_anti-CD95 CH11	1.2	93579_CCD1106 (Keratinocytes)_none	4.2
93103_LAK cells_resting	25.6	93580_CCD1106 (Keratinocytes)_TNFa and IFNg **	4.0
93788_LAK cells_IL-2	3.2	93791_Liver Cirrhosis	1.7
93787_LAK cells_IL-2+IL-12	6.4	93577_NCI-H292	3.3
93789_LAK cells_IL-2+IFN gamma	5.3	93358_NCI-H292_IL-4	5.5
93790_LAK cells_IL-2+ IL-18	4.6	93360_NCI-H292_IL-9	4.6
93104_LAK cells_PMA/ionomycin and IL-18	11.6	93359_NCI-H292_IL-13	6.3
93578_NK Cells IL-2_resting	3.0	93357_NCI-H292_IFN gamma	3.4
93109_Mixed Lymphocyte Reaction_Two Way MLR	17.2	93777_HPAEC_-	0.2
93110_Mixed Lymphocyte Reaction_Two Way MLR	13.5	93778_HPAEC_IL-1 beta/TNA alpha	0.5
93111_Mixed Lymphocyte Reaction_Two Way MLR	5.4	93254_Normal Human Lung Fibroblast_none	0.5
93112_Mononuclear Cells (PBMCs)_resting	9.1	93253_Normal Human Lung Fibroblast_TNFa (4 ng/ml) and	0.3

		IL-1b (1 ng/ml)	
93113_Mononuclear Cells (PBMCs)_PWM	11.5	93257_Normal Human Lung Fibroblast_IL-4	0.2
93114_Mononuclear Cells (PBMCs)_PHA-L	20.5	93256_Normal Human Lung Fibroblast_IL-9	0.3
93249_Ramos (B cell)_none	70.5	93255_Normal Human Lung Fibroblast_IL-13	0.6
93250_Ramos (B cell)_ionomycin	100.0	93258_Normal Human Lung Fibroblast_IFN gamma	0.5
93349_B lymphocytes_PWM	12.4	93106_Dermal Fibroblasts CCD1070_resting	0.5
93350_B lymphoytes_CD40L and IL-4	64.9	93361_Dermal Fibroblasts CCD1070_TNF alpha 4 ng/ml	4.1
92665_EOL-1 (Eosinophil)_dbcAMP differentiated	10.0	93105_Dermal Fibroblasts CCD1070_IL-1 beta 1 ng/ml	0.1
93248_EOL-1 (Eosinophil)_dbcAMP/PMAionomycin	6.9	93772_dermal fibroblast_IFN gamma	1.0
93356_Dendritic Cells_none	10.1	93771_dermal fibroblast_IL-4	2.2
93355_Dendritic Cells_LPS 100 ng/ml	21.1	93892_Dermal fibroblasts_none	1.0
93775_Dendritic Cells_anti-CD40	9.9	99202_Neutrophils_TNFa+LPS	9.2
93774_Monocytes_resting	38.2	99203_Neutrophils_none	38.1
93776_Monocytes_LPS 50 ng/ml	51.1	735010_Colon_normal	1.2
93581_Macrophages_resting	18.4	735019_Lung_none	5.5
93582_Macrophages_LPS 100 ng/ml	19.7	64028-1_Thymus_none	8.2
93098_HUVEC (Endothelial)_none	0.0	64030-1_Kidney_none	9.8
93099_HUVEC (Endothelial)_starved	0.3		

Table 27. Panel CNSD.01

Tissue Name	Relative Expression(%) cns1x4tm6179f_ag1493_b1	Tissue Name	Relative Expression(%) cns1x4tm6179f_ag1493_b1
102633_BA4 Control	14.8	102605_BA17 PSP	15.9
102641_BA4 Control2	51.1	102612_BA17 PSP2	8.3
102625_BA4 Alzheimer's2	3.6	102637_Sub Nigra Control	28.5
102649_BA4 Parkinson's	36.3	102645_Sub Nigra Control2	47.1
102656_BA4 Parkinson's2	54.3	102629_Sub Nigra Alzheimer's2	3.3
102664_BA4 Huntington's	17.2	102660_Sub Nigra Parkinson's2	61.6
102671_BA4 Huntington's2	4.3	102667_Sub Nigra Huntington's	37.6

102603_BA4 PSP	0.0	102674_Sub Nigra Huntington's2	8.2
102610_BA4 PSP2	14.0	102614_Sub Nigra PSP2	4.6
102588_BA4 Depression	14.9	102592_Sub Nigra Depression	6.8
102596_BA4 Depression2	2.6	102599_Sub Nigra Depression2	6.6
102634_BA7 Control	16.5	102636_Glob Palladus Control	10.2
102642_BA7 Control2	13.7	102644_Glob Palladus Control2	9.0
102626_BA7 Alzheimer's2	3.9	102620_Glob Palladus Alzheimer's	4.4
102650_BA7 Parkinson's	18.2	102628_Glob Palladus Alzheimer's2	2.6
102657_BA7 Parkinson's2	45.6	102652_Glob Palladus Parkinson's	80.4
102665_BA7 Huntington's	35.9	102659_Glob Palladus Parkinson's2	5.2
102672_BA7 Huntington's2	18.7	102606_Glob Palladus PSP	6.4
102604_BA7 PSP	54.8	102613_Glob Palladus PSP2	2.7
102611_BA7 PSP2	17.8	102591_Glob Palladus Depression	4.0
102589_BA7 Depression	11.2	102638_Temp Pole Control	13.8
102632_BA9 Control	16.1	102646_Temp Pole Control2	**100.0**
102640_BA9 Control2	96.5	102622_Temp Pole Alzheimer's	0.8
102617_BA9 Alzheimer's	0.0	102630_Temp Pole Alzheimer's2	3.0
102624_BA9 Alzheimer's2	7.0	102653_Temp Pole Parkinson's	51.3
102648_BA9 Parkinson's	29.5	102661_Temp Pole Parkinson's2	34.2
102655_BA9 Parkinson's2	47.8	102668_Temp Pole Huntington's	50.4
102663_BA9 Huntington's	27.6	102607_Temp Pole PSP	1.2
102670_BA9 Huntington's2	14.9	102615_Temp Pole PSP2	1.7
102602_BA9 PSP	9.0	102600_Temp Pole Depression2	8.9
102609_BA9 PSP2	3.1	102639_Cing Gyr Control	50.1
102587_BA9 Depression	3.2	102647_Cing Gyr Control2	55.7
102595_BA9 Depression2	5.3	102623_Cing Gyr Alzheimer's	18.9
102635_BA17 Control	17.9	102631_Cing Gyr Alzheimer's2	0.9
102643_BA17 Control2	37.2	102654_Cing Gyr Parkinson's	35.5
102627_BA17 Alzheimer's2	5.7	102662_Cing Gyr Parkinson's2	84.5
102651_BA17 Parkinson's	45.7	102669_Cing Gyr Huntington's	67.5
102658_BA17 Parkinson's2	18.1	102676_Cing Gyr Huntington's2	23.3
102666_BA17 Huntington's	15.1	102608_Cing Gyr PSP	14.0
102673_BA17 Huntington's2	13.0	102616_Cing Gyr PSP2	7.9
102590_BA17 Depression	9.4	102594_Cing Gyr Depression	3.0
102597_BA17 Depression2	31.9	102601_Cing Gyr Depression2	11.8

Panel 1.2 Summary: Ag1493 The high expression of the MOL3a gene seen in adipose (CT value = 25) is most likely skewed due to genomic DNA contamination in this sample. Otherwise, the gene is expressed mainly in normal tissues, including brain (particularly cerebral cortex), kidney, and prostate. Expression of the MOL3a gene in skeletal muscle and liver may suggest function in metabolic diseases, including obesity and diabetes. Furthermore, MOL3a expression is down regulated in a number of tumor cell lines relative to the normal controls suggesting a potential utility of this gene in the treatment of cancer.

Panel 1.3D Summary: Ag1493 In this panel, highest expression of the MOL3a gene is detected in the amygdala of the brain (CT value = 29.6). This may suggest that the MOL3a gene plays a role in normal brain function, including fear and anxiety response. In addition, high expression is also observed in adipose and bone marrow suggesting potential roles in metabolic and immune function. Overall, expression of the MOL3a gene in panel 1.3D reveals that it is associated mostly with normal tissues. In a couple of instances, the expression of this gene is seen in clusters of cell lines, specifically in breast and ovarian cancer cell lines. Thus, therapeutic modulation of expression of this gene may be of utility in the treatment breast and ovarian cancers. Alternatively, replacement of the MOL3a protein that is missing from some cancer cells using recombinant protein might provide a useful treatment for these types of cancers.

Panel 2D Summary: Ag1493 Expression of the MOL3a gene is highest in thyroid and appears to be widespread across many samples on Panel 2D. However, overall there appears to be generally higher expression in normal tissues when compared to cancerous counterparts. Thus, therapeutic modulation of this gene or gene product might show utility for a range of oncology indications. Semaphorins and their receptors are known signals for axon guidance; they are also suspected to regulate developmental processes involving cell migration and morphogenesis, and have been implicated in immune function and tumor progression.

Panel 4.1D Summary: Ag1493 The MOL3a transcript is highly expressed in a B cell line as well as in B cells stimulated with CD40L and IL4. Expression of this transcript is also found to a lesser degree in monocytes and macrophages independently of their activation status. Of interest, CD100, which is an activation molecule on T cells, is a member of the semaphorin protein family. The semaphorin B-like protein encoded by the MOL3a transcript could therefore also serve as a B cell activation marker. The semaphorin family has additionally been reported to play a role in chemotaxis. Thus, protein therapeutics or monoclonal antibodies raised against the MOL3a protein, could inhibit spontaneous and chemokine induced migration of B cells and monocytes and potentially regulate B cell

differentiation and B cell isotype switching. Regulation of this molecule by protein therapeutics or monoclonal antibodies could also function to regulate immunity and be important for the treatment of autoimmune diseases, allergic diseases, and immune rejection in transplantation. In support of this hypothesis, recent studies indicate that semaphorins bind with high affinity to at least two different receptor families and are biologically active on immune cells as well as neuronal cells (Curr Opin Immunol 1999 Aug;11(4):387-91).

Panel CNSD.01 Summary: Ag1493 Semaphorins can act as axon guidance proteins, specifically through their ability to act as chemorepellents that inhibit CNS regenerative capacity. Although there is considerable variance between individuals in MOL3a gene expression levels in this panel, levels of this protein are reduced to less than 1/3 of that seen in controls in the temporal cortex of Alzheimer's patients (which shows marked synaptogenic loss in mid to late phases of the disease) as well as in diseases not associated with neurodegeneration of the temporal cortex. Therefore, manipulation of levels of this protein may be of use in inducing a compensatory synaptogenic response to neuronal death in Alzheimer's disease.

D. MOL4

Expression of gene MOL4 was assessed using the primer-probe set Ag1216, described in Table 28. Results of the RTQ-PCR runs are shown in Tables 29, 30, 31, and 32.

Table 28. Probe Name: Ag1216

Primers	Sequences	TM	Length	Start Position	SEQ ID NO:
Forward	5'-CCCGAAGAATGAAAAGTACACA-3'	59.1	22	5263	77
Probe	FAM-5'-CCCATGGAATTCAAGACCCTGAACAA-3'-TAMRA	69.7	26	5285	78
Reverse	5'-AATGGGTAGAAGTTGGCTCTGT-3'	59.2	22	5331	79

Table 29. Panel 1.2

Tissue Name	Relative Expression(%) 1.2tm1404f_ag1216	Tissue Name	Relative Expression(%) 1.2tm1404f_ag1216
Endothelial cells	0.0	Renal ca. 786-0	100.0

Endothelial cells (treated)	76.8	Renal ca. A498	3.0
Pancreas	11.3	Renal ca. RXF 393	45.4
Pancreatic ca. CAPAN 2	0.0	Renal ca. ACHN	0.2
Adrenal Gland (new lot*)	12.2	Renal ca. UO-31	0.0
Thyroid	11.6	Renal ca. TK-10	0.0
Salivary gland	2.7	Liver	7.5
Pituitary gland	12.6	Liver (fetal)	7.7
Brain (fetal)	77.9	Liver ca. (hepatoblast) HepG2	0.0
Brain (whole)	73.2	Lung	6.1
Brain (amygdala)	33.7	Lung (fetal)	13.1
Brain (cerebellum)	6.8	Lung ca. (small cell) LX-1	0.0
Brain (hippocampus)	75.3	Lung ca. (small cell) NCI-H69	0.0
Brain (thalamus)	9.9	Lung ca. (s.cell var.) SHP-77	0.0
Cerebral Cortex	87.7	Lung ca. (large cell)NCI-H460	0.0
Spinal cord	8.3	Lung ca. (non-sm. cell) A549	0.0
CNS ca. (glio/astro) U87-MG	0.0	Lung ca. (non-s.cell) NCI-H23	0.0
CNS ca. (glio/astro) U-118-MG	1.5	Lung ca (non-s.cell) HOP-62	0.0
CNS ca. (astro) SW1783	0.0	Lung ca. (non-s.cl) NCI-H522	0.0
CNS ca.* (neuro; met) SK-N-AS	11.1	Lung ca. (squam.) SW 900	0.0
CNS ca. (astro) SF-539	0.0	Lung ca. (squam.) NCI-H596	0.0
CNS ca. (astro) SNB-75	6.4	Mammary gland	13.0
CNS ca. (glio) SNB-19	4.4	Breast ca.* (pl. effusion) MCF-7	0.0
CNS ca. (glio) U251	4.2	Breast ca.* (pl.ef) MDA-MB-231	0.0
CNS ca. (glio) SF-295	0.0	Breast ca.* (pl. effusion) T47D	0.0
Heart	61.1	Breast ca. BT-549	1.8
Skeletal Muscle (new lot*)	8.2	Breast ca. MDA-N	0.0
Bone marrow	0.1	Ovary	10.2
Thymus	0.2	Ovarian ca. OVCAR-3	66.4
Spleen	0.0	Ovarian ca. OVCAR-4	1.3
Lymph node	5.1	Ovarian ca. OVCAR-5	3.5
Colorectal	0.2	Ovarian ca. OVCAR-8	0.0
Stomach	5.9	Ovarian ca. IGROV-1	0.3
Small intestine	11.8	Ovarian ca.* (ascites) SK-OV-3	0.0
Colon ca. SW480	0.0	Uterus	7.7
Colon ca.* (SW480 met)SW620	2.7	Placenta	8.1
Colon ca. HT29	0.0	Prostate	4.6
Colon ca. HCT-116	0.0	Prostate ca.* (bone met)PC-3	0.0
Colon ca. CaCo-2	0.5	Testis	4.6
83219 CC Well to Mod Diff (ODO3866)	0.5	Melanoma Hs688(A).T	0.0
Colon ca. HCC-2998	0.0	Melanoma* (met) Hs688(B).T	0.0
Gastric ca.* (liver met) NCI-N87	0.0	Melanoma UACC-62	0.0
Bladder	6.7	Melanoma M14	0.0

Trachea	3.1	Melanoma LOX IMVI	0.0
Kidney	32.3	Melanoma* (met) SK-MEL-5	0.0
Kidney (fetal)	51.8	Adipose	1.5

Table 30. Panel 2.2

Tissue Name	Relative Expression(%) 2.2x4tm6515f_ag1216_b1	Tissue Name	Relative Expression(%) 2.2x4tm6515f_ag1216_b1
Normal Colon GENPAK 061003	0.1	83793 Kidney NAT (OD04348)	11.6
97759 Colon cancer (OD06064)	0.0	98938 Kidney malignant cancer (OD06204B)	0.0
97760 Colon cancer NAT (OD06064)	0.4	98939 Kidney normal adjacent tissue (OD06204E)	1.4
97778 Colon cancer (OD06159)	0.0	85973 Kidney Cancer (OD04450-01)	1.3
97779 Colon cancer NAT (OD06159)	0.8	85974 Kidney NAT (OD04450-03)	3.2
98861 Colon cancer (OD06297-04)	0.0	Kidney Cancer Clontech 8120613	0.0
98862 Colon cancer NAT (OD06297-015)	0.6	Kidney NAT Clontech 8120614	3.5
83237 CC Gr.2 ascend colon (ODO3921)	0.2	Kidney Cancer Clontech 9010320	3.5
83238 CC NAT (ODO3921)	0.0	Kidney NAT Clontech 9010321	0.9
97766 Colon cancer metastasis (OD06104)	0.2	Kidney Cancer Clontech 8120607	8.3
97767 Lung NAT (OD06104)	1.1	Kidney NAT Clontech 8120608	0.7
87472 Colon mets to lung (OD04451-01)	0.3	Normal Uterus GENPAK 061018	2.9
87473 Lung NAT (OD04451-02)	0.4	Uterus Cancer GENPAK 064011	0.2
Normal Prostate Clontech A+ 6546-1 (8090438)	0.5	Normal Thyroid Clontech A+ 6570-1 (7080817)	0.4
84140 Prostate Cancer (OD04410)	0.2	Thyroid Cancer GENPAK 064010	0.2
84141 Prostate NAT (OD04410)	0.0	Thyroid Cancer INVITROGEN A302152	0.3
Normal Ovary Res. Gen.	2.3	Thyroid NAT INVITROGEN A302153	0.3
98863 Ovarian cancer (OD06283-03)	4.6	Normal Breast GENPAK 061019	1.3
98865 Ovarian cancer NAT/fallopian tube (OD06283-07)	1.3	84877 Breast Cancer (OD04566)	0.4
Ovarian Cancer GENPAK 064008	9.7	Breast Cancer Res. Gen. 1024	0.4
97773 Ovarian cancer (OD06145)	0.0	85975 Breast Cancer (OD04590-01)	1.6

97775 Ovarian cancer NAT (OD06145)	0.0	85976 Breast Cancer Mets (OD04590-03)	0.8
98853 Ovarian cancer (OD06455-03)	0.2	87070 Breast Cancer Metastasis (OD04655-05)	0.6
98854 Ovarian NAT (OD06455-07) Fallopian tube	1.9	GENPAK Breast Cancer 064006	0.9
Normal Lung GENPAK 061010	0.6	Breast Cancer Clontech 9100266	1.4
92337 Invasive poor diff. lung adeno (ODO4945-01	0.4	Breast NAT Clontech 9100265	0.6
92338 Lung NAT (ODO4945-03)	0.2	Breast Cancer INVITROGEN A209073	0.0
84136 Lung Malignant Cancer (OD03126)	0.0	Breast NAT INVITROGEN A2090734	1.4
84137 Lung NAT (OD03126)	0.4	97763 Breast cancer (OD06083)	0.8
90372 Lung Cancer (OD05014A)	1.0	97764 Breast cancer node metastasis (OD06083)	0.9
90373 Lung NAT (OD05014B)	2.1	Normal Liver GENPAK 061009	0.3
97761 Lung cancer (OD06081)	0.8	Liver Cancer Research Genetics RNA 1026	0.7
97762 Lung cancer NAT (OD06081)	0.4	Liver Cancer Research Genetics RNA 1025	1.6
85950 Lung Cancer (OD04237-01)	0.4	Paired Liver Cancer Tissue Research Genetics RNA 6004-T	0.0
85970 Lung NAT (OD04237-02)	0.0	Paired Liver Tissue Research Genetics RNA 6004-N	0.0
83255 Ocular Mel Met to Liver (ODO4310)	0.2	Paired Liver Cancer Tissue Research Genetics RNA 6005-T	1.3
83256 Liver NAT (ODO4310)	0.0	Paired Liver Tissue Research Genetics RNA 6005-N	0.4
84139 Melanoma Mets to Lung (OD04321)	0.7	Liver Cancer GENPAK 064003	0.5
84138 Lung NAT (OD04321)	0.0	Normal Bladder GENPAK 061001	0.8
Normal Kidney GENPAK 061008	1.7	Bladder Cancer Research Genetics RNA 1023	0.7
83786 Kidney Ca, Nuclear grade 2 (OD04338)	5.3	Bladder Cancer INVITROGEN A302173	0.2
83787 Kidney NAT (OD04338)	6.2	Normal Stomach GENPAK 061017	1.0
83788 Kidney Ca Nuclear grade 1/2 (OD04339)	**100.0**	Gastric Cancer Clontech 9060397	0.2
83789 Kidney NAT (OD04339)	3.0	NAT Stomach Clontech 9060396	0.4
83790 Kidney Ca, Clear cell type (OD04340)	.26.3	Gastric Cancer Clontech 9060395	0.4
83791 Kidney NAT (OD04340)	0.9	NAT Stomach Clontech 9060394	0.7
83792 Kidney Ca, Nuclear	5.1	Gastric Cancer GENPAK	1.1

grade 3 (OD04348)		064005	

Table 31. Panel 4D

Tissue Name	Relative Expression(%)	
	4Dtm2072f_ag1216	4Dtm2246f_ag1216
93768_Secondary Th1_anti-CD28/anti-CD3	0.0	0.0
93769_Secondary Th2_anti-CD28/anti-CD3	0.0	0.0
93770_Secondary Tr1_anti-CD28/anti-CD3	0.0	0.0
93573_Secondary Th1_resting day 4-6 in IL-2	0.0	0.0
93572_Secondary Th2_resting day 4-6 in IL-2	0.0	0.0
93571_Secondary Tr1_resting day 4-6 in IL-2	0.0	0.0
93568_primary Th1_anti-CD28/anti-CD3	0.0	0.0
93569_primary Th2_anti-CD28/anti-CD3	0.0	0.0
93570_primary Tr1_anti-CD28/anti-CD3	0.0	0.0
93565_primary Th1_resting dy 4-6 in IL-2	0.0	0.0
93566_primary Th2_resting dy 4-6 in IL-2	1.1	0.0
93567_primary Tr1_resting dy 4-6 in IL-2	0.0	0.9
93351_CD45RA CD4 lymphocyte_anti-CD28/anti-CD3	0.0	0.0
93352_CD45RO CD4 lymphocyte_anti-CD28/anti-CD3	0.8	0.0
93251_CD8 Lymphocytes_anti-CD28/anti-CD3	0.0	0.0
93353_chronic CD8 Lymphocytes 2ry_resting dy 4-6 in IL-2	0.0	1.3
93574_chronic CD8 Lymphocytes 2ry_activated CD3/CD28	0.0	0.0
93354_CD4_none	0.0	0.9
93252_Secondary Th1/Th2/Tr1_anti-CD95 CH11	0.0	0.0
93103_LAK cells_resting	0.0	0.0
93788_LAK cells_IL-2	0.0	0.0
93787_LAK cells_IL-2+IL-12	0.0	1.9
93789_LAK cells_IL-2+IFN gamma	1.1	0.0
93790_LAK cells_IL-2+ IL-18	0.0	0.0
93104_LAK cells_PMA/ionomycin and IL-18	0.0	0.0
93578_NK Cells IL-2_resting	0.0	0.0
93109_Mixed Lymphocyte Reaction_Two Way MLR	0.0	0.0
93110_Mixed Lymphocyte Reaction_Two Way MLR	0.0	0.0
93111_Mixed Lymphocyte Reaction_Two Way MLR	0.0	0.0
93112_Mononuclear Cells (PBMCs)_resting	0.0	0.0
93113_Mononuclear Cells (PBMCs)_PWM	2.8	2.5
93114_Mononuclear Cells (PBMCs)_PHA-L	0.0	3.1
93249_Ramos (B cell)_none	0.0	0.0
93250_Ramos (B cell)_ionomycin	0.0	0.0
93349_B lymphocytes_PWM	0.0	0.0
93350_B lymphoytes_CD40L and IL-4	1.1	0.0
92665_EOL-1 (Eosinophil)_dbcAMP differentiated	0.0	0.0
93248_EOL-1 (Eosinophil)_dbcAMP/PMAionomycin	0.0	0.0

93356_Dendritic Cells_none	0.0	0.0
93355_Dendritic Cells_LPS 100 ng/ml	0.0	0.0
93775_Dendritic Cells_anti-CD40	0.0	0.0
93774_Monocytes_resting	0.0	0.0
93776_Monocytes_LPS 50 ng/ml	0.0	0.0
93581_Macrophages_resting	0.0	0.0
93582_Macrophages_LPS 100 ng/ml	0.8	0.0
93098_HUVEC (Endothelial)_none	0.0	0.0
93099_HUVEC (Endothelial)_starved	0.0	0.0
93100_HUVEC (Endothelial)_IL-1b	0.0	1.3
93779_HUVEC (Endothelial)_IFN gamma	0.0	0.0
93102_HUVEC (Endothelial)_TNF alpha + IFN gamma	0.0	0.0
93101_HUVEC (Endothelial)_TNF alpha + IL4	0.0	0.0
93781_HUVEC (Endothelial)_IL-11	0.9	0.0
93583_Lung Microvascular Endothelial Cells_none	0.0	0.0
93584_Lung Microvascular Endothelial Cells_TNFa (4 ng/ml) and IL1b (1 ng/ml)	0.0	0.0
92662_Microvascular Dermal endothelium_none	0.0	0.0
92663_Microsvasular Dermal endothelium_TNFa (4 ng/ml) and IL1b (1 ng/ml)	0.0	0.0
93773_Bronchial epithelium_TNFa (4 ng/ml) and IL1b (1 ng/ml) **	20.3	31.9
93347_Small Airway Epithelium_none	1.8	2.0
93348_Small Airway Epithelium_TNFa (4 ng/ml) and IL1b (1 ng/ml)	2.7	2.5
92668_Coronery Artery SMC_resting	0.0	0.0
92669_Coronery Artery SMC_TNFa (4 ng/ml) and IL1b (1 ng/ml)	0.0	0.0
93107_astrocytes_resting	12.8	16.3
93108_astrocytes_TNFa (4 ng/ml) and IL1b (1 ng/ml)	5.3	13.5
92666_KU-812 (Basophil)_resting	0.0	0.0
92667_KU-812 (Basophil)_PMA/ionoycin	0.0	0.0
93579_CCD1106 (Keratinocytes)_none	1.4	0.0
93580_CCD1106 (Keratinocytes)_TNFa and IFNg **	0.0	4.1
93791_Liver Cirrhosis	1.9	2.3
93792_Lupus Kidney	5.9	10.3
93577_NCI-H292	1.6	0.0
93358_NCI-H292_IL-4	0.7	0.0
93360_NCI-H292_IL-9	0.0	2.8
93359_NCI-H292_IL-13	0.0	0.0
93357_NCI-H292_IFN gamma	0.0	0.0
93777_HPAEC_-	0.0	0.0
93778_HPAEC_IL-1 beta/TNA alpha	0.0	0.0
93254_Normal Human Lung Fibroblast_none	0.0	0.0
93253_Normal Human Lung Fibroblast_TNFa (4 ng/ml) and IL-1b (1 ng/ml)	0.0	0.0
93257_Normal Human Lung Fibroblast_IL-4	0.0	0.0

93256_Normal Human Lung Fibroblast_IL-9	0.0	0.0
93255_Normal Human Lung Fibroblast_IL-13	0.0	0.0
93258_Normal Human Lung Fibroblast_IFN gamma	0.0	0.0
93106_Dermal Fibroblasts CCD1070_resting	0.0	0.0
93361_Dermal Fibroblasts CCD1070_TNF alpha 4 ng/ml	0.0	0.0
93105_Dermal Fibroblasts CCD1070_IL-1 beta 1 ng/ml	0.0	0.0
93772_dermal fibroblast_IFN gamma	0.0	0.0
93771_dermal fibroblast_IL-4	0.0	1.5
93259_IBD Colitis 1**	1.5	2.8
93260_IBD Colitis 2	0.0	1.2
93261_IBD Crohns	1.4	4.0
735010_Colon_normal	3.4	2.7
735019_Lung_none	52.1	42.3
64028-1_Thymus_none	100.0	100.0
64030-1_Kidney_none	1.6	0.0

Panel 32. Panel CNSD.01

Tissue Name	Relative Expression(%) cns1x4tm6177fag1216_b2	Tissue Name	Relative Expression(%) cns1x4tm6177fag1216_b2
102633_BA4 Control	18.7	102605_BA17 PSP	38.3
102641_BA4 Control2	62.0	102612_BA17 PSP2	19.2
102625_BA4 Alzheimer's2	17.2	102637_Sub Nigra Control	6.8
102649_BA4 Parkinson's	79.4	102645_Sub Nigra Control2	12.2
102656_BA4 Parkinson's2	54.7	102629_Sub Nigra Alzheimer's2	10.8
102664_BA4 Huntington's	37.5	102660_Sub Nigra Parkinson's2	17.4
102671_BA4 Huntington's2	17.0	102667_Sub Nigra Huntington's	21.1
102603_BA4 PSP	16.5	102674_Sub Nigra Huntington's2	8.6
102610_BA4 PSP2	22.9	102614_Sub Nigra PSP2	1.9
102588_BA4 Depression	45.0	102592_Sub Nigra Depression	1.9
102596_BA4 Depression2	27.0	102599_Sub Nigra Depression2	3.6
102634_BA7 Control	45.2	102636_Glob Palladus Control	21.2
102642_BA7 Control2	25.5	102644_Glob Palladus Control2	12.8
102626_BA7 Alzheimer's2	20.0	102620_Glob Palladus Alzheimer's	3.0
102650_BA7 Parkinson's	48.1	102628_Glob Palladus Alzheimer's2	4.2
102657_BA7 Parkinson's2	27.3	102652_Glob Palladus Parkinson's	100.0
102665_BA7 Huntington's	58.5	102659_Glob Palladus Parkinson's2	10.6
102672_BA7 Huntington's2	65.1	102606_Glob Palladus PSP	4.3

102604_BA7 PSP	93.8	102613_Glob Palladus PSP2	6.8
102611_BA7 PSP2	34.8	102591_Glob Palladus Depression	9.1
102589_BA7 Depression	20.9	102638_Temp Pole Control	29.7
102632_BA9 Control	35.4	102646_Temp Pole Control2	63.9
102640_BA9 Control2	58.7	102622_Temp Pole Alzheimer's	7.3
102617_BA9 Alzheimer's	9.2	102630_Temp Pole Alzheimer's2	16.4
102624_BA9 Alzheimer's2	36.1	102653_Temp Pole Parkinson's	88.4
102648_BA9 Parkinson's	33.0	102661_Temp Pole Parkinson's2	50.2
102655_BA9 Parkinson's2	32.2	102668_Temp Pole Huntington's	63.5
102663_BA9 Huntington's	50.4	102607_Temp Pole PSP	22.1
102670_BA9 Huntington's2	18.6	102615_Temp Pole PSP2	11.4
102602_BA9 PSP	14.3	102600_Temp Pole Depression2	36.9
102609_BA9 PSP2	7.6	102639_Cing Gyr Control	60.1
102587_BA9 Depression	14.9	102647_Cing Gyr Control2	48.0
102595_BA9 Depression2	18.3	102623_Cing Gyr Alzheimer's	19.4
102635_BA17 Control	87.9	102631_Cing Gyr Alzheimer's2	24.8
102643_BA17 Control2	73.0	102654_Cing Gyr Parkinson's	29.7
102627_BA17 Alzheimer's2	36.7	102662_Cing Gyr Parkinson's2	33.2
102651_BA17 Parkinson's	81.0	102669_Cing Gyr Huntington's	36.0
102658_BA17 Parkinson's2	95.4	102676_Cing Gyr Huntington's2	27.7
102666_BA17 Huntington's	78.7	102608_Cing Gyr PSP	20.7
102673_BA17 Huntington's2	37.6	102616_Cing Gyr PSP2	3.9
102590_BA17 Depression	33.7	102594_Cing Gyr Depression	34.6
102597_BA17 Depression2	73.7	102601_Cing Gyr Depression2	21.0

Panel 1.2 Summary: Ag1216 The MOL4 gene is well expressed in a variety of normal tissues including kidney, heart, brain, thymus and lung. Of interest is the robust expression in activated endothelial cells, which may indicate that this gene is important for angiogenesis or lymphocyte trafficking. Inflammatory lymphocytes preferentially traffic into tissues by crossing activated endothelium. Expression of the MOL4 gene appears to be up regulated in renal cell carcinomas. In contrast, expression of the MOL4 gene is down regulated in a number of cancer cell lines (including pancreatic, CNS, breast, and lung) relative to the normal controls. No expression of this gene is detected in a variety of melanoma cell lines. Therefore, modulation of MOL4 gene function may provide an effective treatment for a variety of cancers.

Panel 2.2 Summary: Ag1216 Expression of the MOL4 gene appears to be associated with kidney cancers. This is in good agreement with the data obtained in Panel 1.2 and

suggests that therapeutic modulation of this gene using inhibitory monoclonal antibodies or small molecules may prove useful in the treatment of kidney cancers. In addition, the MOL4 gene may be a useful marker for the detection of renal cell carcinomas.

Panel 4D Summary: Ag1216 Two replicate experiments using the same probe and primer set were in good agreement. The MOL4 transcript is highly expressed in thymus. To a much lesser degree, the transcript is also expressed in the lung as well as in small airway epithelium treated with TNF-a and IL-1b. Therefore, protein therapeutics designed against the protein encoded for by this transcript could reduce inflammation in asthma or other lung disease such as emphysema.

Panel CNSD.01 Summary: Ag1216 Semaphorins can act as axon guidance proteins, specifically through their ability to act as chemorepellents that inhibit CNS regenerative capacity. Manipulation of levels of the MOL4 semaphorin-like protein may therefore be of use in inducing a compensatory synaptogenic response to neuronal death in Alzheimer's disease, Parkinson's disease, Huntington's disease, spinocerebellar ataxia, progressive supranuclear palsy, multiple sclerosis, ALS, head trauma, stroke, or any other disease/condition associated with neuronal loss.

E. MOL5a

Expression of gene MOL5a was assessed using the primer-probe sets Ag1215 and Ag1382 (identical sequences), described in Tables 33 and 34. Results of the RTQ-PCR runs are shown in Tables 35, 36, and 37.

Table 33. Probe Name Ag1215/Ag1382

Primers	Sequences	TM	Length	Start Position	SEQ ID NO:
Forward	5'-AACCCATTATCCTGCGTAACAT-3'	59.6	22	619	80
Probe	FAM-5'-CCCCACCACTCCATGAAGACAGAGTA-3'-TAMRA	68.5	26	645	81
Reverse	5'-CCTACAAAGTGAGGTTCGTTGA-3'	59.3	22	685	82

Table 34. Panel 1.2

Tissue Name	Relative Expression(%)	Relative Expression(%)

	1.2tm1403f_ag1215	1.2tm1581t_ag1382
Endothelial cells	21.9	17.3
Endothelial cells (treated)	100.0	40.9
Pancreas	14.7	0.4
Pancreatic ca. CAPAN 2	3.4	0.7
Adrenal Gland (new lot*)	26.4	10.4
Thyroid	18.8	0.3
Salivary gland	21.5	5.8
Pituitary gland	23.8	1.4
Brain (fetal)	33.9	1.4
Brain (whole)	69.3	3.8
Brain (amygdala)	13.9	3.6
Brain (cerebellum)	67.4	3.0
Brain (hippocampus)	45.1	12.6
Brain (thalamus)	19.1	26.4
Cerebral Cortex	53.2	30.1
Spinal cord	45.4	7.2
CNS ca. (glio/astro) U87-MG	7.8	21.3
CNS ca. (glio/astro) U-118-MG	6.8	13.5
CNS ca. (astro) SW1783	2.0	1.8
CNS ca.* (neuro; met) SK-N-AS	50.3	27.5
CNS ca. (astro) SF-539	20.3	21.2
CNS ca. (astro) SNB-75	13.7	2.8
CNS ca. (glio) SNB-19	19.5	28.9
CNS ca. (glio) U251	15.2	0.0
CNS ca. (glio) SF-295	15.0	5.1
Heart	63.3	12.2
Skeletal Muscle (new lot*)	27.7	0.8
Bone marrow	2.9	0.9
Thymus	4.3	2.8
Spleen	33.4	15.6
Lymph node	23.7	5.4
Colorectal	8.4	2.4
Stomach	36.3	9.6
Small intestine	27.9	11.7
Colon ca. SW480	6.1	1.8
Colon ca.* (SW480 met)SW620	10.4	40.3
Colon ca. HT29	3.0	2.5
Colon ca. HCT-116	6.9	11.0
Colon ca. CaCo-2	13.1	28.9
83219 CC Well to Mod Diff (ODO3866)	3.1	1.8
Colon ca. HCC-2998	12.6	9.5
Gastric ca.* (liver met) NCI-N87	18.4	17.6
Bladder	40.1	15.5

Trachea	13.8	6.0
Kidney	19.5	42.9
Kidney (fetal)	30.4	61.6
Renal ca. 786-0	6.1	6.7
Renal ca. A498	13.0	9.3
Renal ca. RXF 393	8.4	6.0
Renal ca. ACHN	10.7	4.9
Renal ca. UO-31	8.8	3.6
Renal ca. TK-10	15.4	6.2
Liver	15.2	3.2
Liver (fetal)	13.3	5.9
Liver ca. (hepatoblast) HepG2	21.8	12.1
Lung	19.3	0.7
Lung (fetal)	28.1	4.5
Lung ca. (small cell) LX-1	24.5	24.0
Lung ca. (small cell) NCI-H69	8.1	12.7
Lung ca. (s.cell var.) SHP-77	3.8	1.8
Lung ca. (large cell)NCI-H460	40.3	19.5
Lung ca. (non-sm. cell) A549	13.1	13.9
Lung ca. (non-s.cell) NCI-H23	28.1	24.5
Lung ca (non-s.cell) HOP-62	42.3	9.7
Lung ca. (non-s.cl) NCI-H522	90.1	44.1
Lung ca. (squam.) SW 900	37.4	57.8
Lung ca. (squam.) NCI-H596	9.8	4.2
Mammary gland	42.6	9.0
Breast ca.* (pl. effusion) MCF-7	85.3	56.6
Breast ca.* (pl.ef) MDA-MB-231	5.3	1.4
Breast ca.* (pl. effusion) T47D	5.6	16.6
Breast ca. BT-549	5.0	2.6
Breast ca. MDA-N	16.7	3.7
Ovary	49.0	23.0
Ovarian ca. OVCAR-3	46.3	50.7
Ovarian ca. OVCAR-4	11.3	10.6
Ovarian ca. OVCAR-5	28.5	66.9
Ovarian ca. OVCAR-8	19.8	**100.0**
Ovarian ca. IGROV-1	18.2	13.3
Ovarian ca.* (ascites) SK-OV-3	25.7	10.3
Uterus	24.5	3.5
Placenta	90.1	82.9
Prostate	28.5	15.0
Prostate ca.* (bone met)PC-3	39.0	15.4
Testis	8.4	0.6
Melanoma Hs688(A).T	3.8	1.2
Melanoma* (met) Hs688(B).T	2.9	0.8
Melanoma UACC-62	23.2	11.4

Melanoma M14	13.7	6.7
Melanoma LOX IMVI	9.6	1.8
Melanoma* (met) SK-MEL-5	27.4	8.9
Adipose	6.2	12.8

Table 35. Panel 2.2

Tissue Name	Relative Expression(%) 2.2x4tm6515f_ag1215_a2	Tissue Name	Relative Expression(%) 2.2x4tm6515f_ag1215_a2
Normal Colon GENPAK 061003	31.4	83793 Kidney NAT (OD04348)	39.7
97759 Colon cancer (OD06064)	34.3	98938 Kidney malignant cancer (OD06204B)	15.4
97760 Colon cancer NAT (OD06064)	18.6	98939 Kidney normal adjacent tissue (OD06204E)	10.4
97778 Colon cancer (OD06159)	2.7	85973 Kidney Cancer (OD04450-01)	15.3
97779 Colon cancer NAT (OD06159)	22.5	85974 Kidney NAT (OD04450-03)	20.3
98861 Colon cancer (OD06297-04)	3.2	Kidney Cancer Clontech 8120613	1.4
98862 Colon cancer NAT (OD06297-015)	30.6	Kidney NAT Clontech 8120614	20.3
83237 CC Gr.2 ascend colon (ODO3921)	11.6	Kidney Cancer Clontech 9010320	10.5
83238 CC NAT (ODO3921)	8.4	Kidney NAT Clontech 9010321	9.0
97766 Colon cancer metastasis (OD06104)	5.1	Kidney Cancer Clontech 8120607	32.2
97767 Lung NAT (OD06104)	8.3	Kidney NAT Clontech 8120608	12.1
87472 Colon mets to lung (OD04451-01)	15.3	Normal Uterus GENPAK 061018	31.8
87473 Lung NAT (OD04451-02)	4.3	Uterus Cancer GENPAK 064011	31.4
Normal Prostate Clontech A+ 6546-1 (8090438)	11.2	Normal Thyroid Clontech A+ 6570-1 (7080817)	3.0
84140 Prostate Cancer (OD04410)	10.0	Thyroid Cancer GENPAK 064010	14.8
84141 Prostate NAT (OD04410)	14.3	Thyroid Cancer INVITROGEN A302152	37.7
Normal Ovary Res. Gen.	74.7	Thyroid NAT INVITROGEN A302153	7.0
98863 Ovarian cancer (OD06283-03)	27.6	Normal Breast GENPAK 061019	35.4
98865 Ovarian cancer NAT/fallopian tube (OD06283-07)	6.4	84877 Breast Cancer (OD04566)	15.6
Ovarian Cancer GENPAK	16.2	Breast Cancer Res. Gen. 1024	51.4

064008			
97773 Ovarian cancer (OD06145)	8.8	85975 Breast Cancer (OD04590-01)	36.8
97775 Ovarian cancer NAT (OD06145)	24.6	85976 Breast Cancer Mets (OD04590-03)	21.0
98853 Ovarian cancer (OD06455-03)	9.9	87070 Breast Cancer Metastasis (OD04655-05)	66.6
98854 Ovarian NAT (OD06455-07) Fallopian tube	12.9	GENPAK Breast Cancer 064006	19.0
Normal Lung GENPAK 061010	18.0	Breast Cancer Clontech 9100266	27.7
92337 Invasive poor diff. lung adeno (ODO4945-01	11.8	Breast NAT Clontech 9100265	21.6
92338 Lung NAT (ODO4945-03)	13.8	Breast Cancer INVITROGEN A209073	15.1
84136 Lung Malignant Cancer (OD03126)	31.7	Breast NAT INVITROGEN A2090734	28.5
84137 Lung NAT (OD03126)	5.6	97763 Breast cancer (OD06083)	**100.0**
90372 Lung Cancer (OD05014A)	17.7	97764 Breast cancer node metastasis (OD06083)	64.9
90373 Lung NAT (OD05014B)	13.7	Normal Liver GENPAK 061009	17.1
97761 Lung cancer (OD06081)	10.2	Liver Cancer Research Genetics RNA 1026	15.0
97762 Lung cancer NAT (OD06081)	8.2	Liver Cancer Research Genetics RNA 1025	36.3
85950 Lung Cancer (OD04237-01)	15.0	Paired Liver Cancer Tissue Research Genetics RNA 6004-T	18.1
85970 Lung NAT (OD04237-02)	24.0	Paired Liver Tissue Research Genetics RNA 6004-N	6.5
83255 Ocular Mel Met to Liver (ODO4310)	25.5	Paired Liver Cancer Tissue Research Genetics RNA 6005-T	33.3
83256 Liver NAT (ODO4310)	18.1	Paired Liver Tissue Research Genetics RNA 6005-N	31.6
84139 Melanoma Mets to Lung (OD04321)	41.3	Liver Cancer GENPAK 064003	8.9
84138 Lung NAT (OD04321)	9.1	Normal Bladder GENPAK 061001	14.7
Normal Kidney GENPAK 061008	7.5	Bladder Cancer Research Genetics RNA 1023	6.0
83786 Kidney Ca, Nuclear grade 2 (OD04338)	34.7	Bladder Cancer INVITROGEN A302173	28.9
83787 Kidney NAT (OD04338)	7.6	Normal Stomach GENPAK 061017	33.6
83788 Kidney Ca Nuclear grade 1/2 (OD04339)	38.3	Gastric Cancer Clontech 9060397	3.6
83789 Kidney NAT (OD04339)	6.8	NAT Stomach Clontech 9060396	12.2
83790 Kidney Ca, Clear cell type (OD04340)	19.2	Gastric Cancer Clontech 9060395	15.1

83791 Kidney NAT (OD04340)	18.7	NAT Stomach Clontech 9060394	21.2
83792 Kidney Ca, Nuclear grade 3 (OD04348)	10.4	Gastric Cancer GENPAK 064005	17.4

Panel 36. Panel 4D

Tissue Name	Relative Expression(%) 4Dtm2070f_ ag1215	Relative Expression(%) 4Dtm2425t_ ag1382
93768_Secondary Th1_anti-CD28/anti-CD3	27.9	19.6
93769_Secondary Th2_anti-CD28/anti-CD3	35.4	25.5
93770_Secondary Tr1_anti-CD28/anti-CD3	42.0	37.6
93573_Secondary Th1_resting day 4-6 in IL-2	29.5	18.8
93572_Secondary Th2_resting day 4-6 in IL-2	27.5	21.9
93571_Secondary Tr1_resting day 4-6 in IL-2	33.7	23.2
93568_primary Th1_anti-CD28/anti-CD3	35.1	28.1
93569_primary Th2_anti-CD28/anti-CD3	31.4	25.7
93570_primary Tr1_anti-CD28/anti-CD3	55.9	42.6
93565_primary Th1_resting dy 4-6 in IL-2	91.4	**100.0**
93566_primary Th2_resting dy 4-6 in IL-2	68.8	64.6
93567_primary Tr1_resting dy 4-6 in IL-2	55.5	52.1
93351_CD45RA CD4 lymphocyte_anti-CD28/anti-CD3	21.6	17.8
93352_CD45RO CD4 lymphocyte_anti-CD28/anti-CD3	24.5	17.2
93251_CD8 Lymphocytes_anti-CD28/anti-CD3	22.1	15.5
93353_chronic CD8 Lymphocytes 2ry_resting dy 4-6 in IL-2	17.1	12.4
93574_chronic CD8 Lymphocytes 2ry_activated CD3/CD28	29.7	18.2
93354_CD4_none	17.9	15.4
93252_Secondary Th1/Th2/Tr1_anti-CD95 CH11	71.7	67.4
93103_LAK cells_resting	17.1	10.3
93788_LAK cells_IL-2	21.5	15.7
93787_LAK cells_IL-2+IL-12	18.4	17.1
93789_LAK cells_IL-2+IFN gamma	29.5	22.2
93790_LAK cells_IL-2+ IL-18	18.0	22.1
93104_LAK cells_PMA/ionomycin and IL-18	10.7	7.3
93578_NK Cells IL-2_resting	31.9	17.2
93109_Mixed Lymphocyte Reaction_Two Way MLR	9.5	7.6
93110_Mixed Lymphocyte Reaction_Two Way MLR	6.2	6.8
93111_Mixed Lymphocyte Reaction_Two Way MLR	10.5	8.5
93112_Mononuclear Cells (PBMCs)_resting	10.7	10.3
93113_Mononuclear Cells (PBMCs)_PWM	50.7	38.4
93114_Mononuclear Cells (PBMCs)_PHA-L	32.5	31.2
93249_Ramos (B cell)_none	0.0	0.0

93250_Ramos (B cell)_ionomycin	0.0	0.0
93349_B lymphocytes_PWM	36.1	34.2
93350_B lymphoytes_CD40L and IL-4	18.3	17.2
92665_EOL-1 (Eosinophil)_dbcAMP differentiated	43.8	30.1
93248_EOL-1 (Eosinophil)_dbcAMP/PMAionomycin	73.2	63.7
93356_Dendritic Cells_none	1.6	0.6
93355_Dendritic Cells_LPS 100 ng/ml	1.2	5.0
93775_Dendritic Cells_anti-CD40	0.9	0.9
93774_Monocytes_resting	2.9	1.8
93776_Monocytes_LPS 50 ng/ml	29.7	19.2
93581_Macrophages_resting	4.9	2.3
93582_Macrophages_LPS 100 ng/ml	7.2	4.4
93098_HUVEC (Endothelial)_none	9.2	6.5
93099_HUVEC (Endothelial)_starved	18.6	14.7
93100_HUVEC (Endothelial)_IL-1b	3.9	2.6
93779_HUVEC (Endothelial)_IFN gamma	19.2	17.1
93102_HUVEC (Endothelial)_TNF alpha + IFN gamma	2.1	2.4
93101_HUVEC (Endothelial)_TNF alpha + IL4	15.3	12.2
93781_HUVEC (Endothelial)_IL-11	13.6	15.3
93583_Lung Microvascular Endothelial Cells_none	19.9	20.4
93584_Lung Microvascular Endothelial Cells_TNFa (4 ng/ml) and IL1b (1 ng/ml)	18.0	14.9
92662_Microvascular Dermal endothelium_none	29.7	26.8
92663_Microsvasular Dermal endothelium_TNFa (4 ng/ml) and IL1b (1 ng/ml)	36.9	33.4
93773_Bronchial epithelium_TNFa (4 ng/ml) and IL1b (1 ng/ml) **	41.5	25.7
93347_Small Airway Epithelium_none	13.3	8.8
93348_Small Airway Epithelium_TNFa (4 ng/ml) and IL1b (1 ng/ml)	56.6	45.4
92668_Coronery Artery SMC_resting	22.2	19.8
92669_Coronery Artery SMC_TNFa (4 ng/ml) and IL1b (1 ng/ml)	28.1	19.9
93107_astrocytes_resting	13.0	20.2
93108_astrocytes_TNFa (4 ng/ml) and IL1b (1 ng/ml)	23.7	18.6
92666_KU-812 (Basophil)_resting	97.3	75.3
92667_KU-812 (Basophil)_PMA/ionoycin	**100.0**	90.1
93579_CCD1106 (Keratinocytes)_none	16.6	13.8
93580_CCD1106 (Keratinocytes)_TNFa and IFNg **	67.8	11.3
93791_Liver Cirrhosis	14.3	12.3
93792_Lupus Kidney	28.7	18.8
93577_NCI-H292	27.4	28.3
93358_NCI-H292_IL-4	61.1	55.1
93360_NCI-H292_IL-9	37.6	31.6
93359_NCI-H292_IL-13	44.4	42.9
93357_NCI-H292_IFN gamma	21.0	20.3
93777_HPAEC_-	21.5	17.2

93778_HPAEC_IL-1 beta/TNA alpha	11.0	9.5
93254_Normal Human Lung Fibroblast_none	31.0	20.6
93253_Normal Human Lung Fibroblast_TNFa (4 ng/ml) and IL-1b (1 ng/ml)	26.1	23.5
93257_Normal Human Lung Fibroblast_IL-4	40.6	37.1
93256_Normal Human Lung Fibroblast_IL-9	21.3	17.7
93255_Normal Human Lung Fibroblast_IL-13	56.3	53.2
93258_Normal Human Lung Fibroblast_IFN gamma	59.5	45.4
93106_Dermal Fibroblasts CCD1070_resting	47.0	33.4
93361_Dermal Fibroblasts CCD1070_TNF alpha 4 ng/ml	73.7	55.1
93105_Dermal Fibroblasts CCD1070_IL-1 beta 1 ng/ml	36.6	37.4
93772_dermal fibroblast_IFN gamma	8.8	9.7
93771_dermal fibroblast_IL-4	20.2	19.6
93259_IBD Colitis 1**	13.2	10.7
93260_IBD Colitis 2	4.1	2.0
93261_IBD Crohns	3.1	3.1
735010_Colon_normal	29.9	24.3
735019_Lung_none	46.0	38.4
64028-1_Thymus_none	60.3	54.0
64030-1_Kidney_none	32.1	26.4

Table 37. Panel CNSD.01

Tissue Name	Relative Expression(%) cns1x4tm6177fag1215_b1	Tissue Name	Relative Expression(%) cns1x4tm6177fag1215_b1
102633_BA4 Control	20.2	102605_BA17 PSP	31.1
102641_BA4 Control2	34.5	102612_BA17 PSP2	9.9
102625_BA4 Alzheimer's2	9.0	102637_Sub Nigra Control	54.5
102649_BA4 Parkinson's	32.9	102645_Sub Nigra Control2	31.7
102656_BA4 Parkinson's2	46.8	102629_Sub Nigra Alzheimer's2	24.2
102664_BA4 Huntington's	36.2	102660_Sub Nigra Parkinson's2	91.5
102671_BA4 Huntington's2	11.1	102667_Sub Nigra Huntington's	80.6
102603_BA4 PSP	15.8	102674_Sub Nigra Huntington's2	43.0
102610_BA4 PSP2	45.9	102614_Sub Nigra PSP2	24.8
102588_BA4 Depression	19.6	102592_Sub Nigra Depression	18.5
102596_BA4 Depression2	7.8	102599_Sub Nigra Depression2	12.0
102634_BA7 Control	26.1	102636_Glob Palladus Control	20.2
102642_BA7 Control2	32.8	102644_Glob Palladus Control2	10.7
102626_BA7 Alzheimer's2	4.5	102620_Glob Palladus Alzheimer's	21.6
102650_BA7 Parkinson's	22.9	102628_Glob Palladus Alzheimer's2	4.9

102657_BA7 Parkinson's2	35.5	102652_Glob Palladus Parkinson's	100.0
102665_BA7 Huntington's	39.1	102659_Glob Palladus Parkinson's2	23.0
102672_BA7 Huntington's2	40.4	102606_Glob Palladus PSP	10.1
102604_BA7 PSP	31.2	102613_Glob Palladus PSP2	11.8
102611_BA7 PSP2	32.2	102591_Glob Palladus Depression	23.4
102589_BA7 Depression	6.2	102638_Temp Pole Control	7.7
102632_BA9 Control	12.5	102646_Temp Pole Control2	29.2
102640_BA9 Control2	48.5	102622_Temp Pole Alzheimer's	4.0
102617_BA9 Alzheimer's	6.9	102630_Temp Pole Alzheimer's2	3.6
102624_BA9 Alzheimer's2	5.4	102653_Temp Pole Parkinson's	22.5
102648_BA9 Parkinson's	25.2	102661_Temp Pole Parkinson's2	22.1
102655_BA9 Parkinson's2	33.4	102668_Temp Pole Huntington's	30.6
102663_BA9 Huntington's	44.8	102607_Temp Pole PSP	3.8
102670_BA9 Huntington's2	13.7	102615_Temp Pole PSP2	2.2
102602_BA9 PSP	23.4	102600_Temp Pole Depression2	7.1
102609_BA9 PSP2	5.0	102639_Cing Gyr Control	58.9
102587_BA9 Depression	10.8	102647_Cing Gyr Control2	40.2
102595_BA9 Depression2	11.8	102623_Cing Gyr Alzheimer's	19.3
102635_BA17 Control	39.6	102631_Cing Gyr Alzheimer's2	11.8
102643_BA17 Control2	40.0	102654_Cing Gyr Parkinson's	42.1
102627_BA17 Alzheimer's2	7.6	102662_Cing Gyr Parkinson's2	35.7
102651_BA17 Parkinson's	31.2	102669_Cing Gyr Huntington's	62.0
102658_BA17 Parkinson's2	35.2	102676_Cing Gyr Huntington's2	28.7
102666_BA17 Huntington's	47.2	102608_Cing Gyr PSP	66.8
102673_BA17 Huntington's2	25.0	102616_Cing Gyr PSP2	5.2
102590_BA17 Depression	22.2	102594_Cing Gyr Depression	14.7
102597_BA17 Depression2	41.7	102601_Cing Gyr Depression2	32.0

Panel 1.2 Summary: Ag1215/Ag1382 Two replicate experiments were performed using probe and primer sets of identical sequences; however, relatively disparate results were obtained on this panel. For Ag1215, the MOL5a gene is expressed at high levels across most of the tissues on this panel with highest expression in treated endothelial cells (CT value = 23). For Ag1382, the MOL5a gene is expressed at high levels across most of the tissues on this panel with highest expression in an ovarian cancer cell line (CT value = 22). To summarize the expression profile, there appears to be widespread expression of the MOL5a gene in a number of tissues and cell lines. Furthermore, the expression of this gene seems to be associated with reproductive tissues and cancer cell lines whose origins are such. For

instance, there is significant expression in ovarian cell lines, breast cell lines and placenta tissue. There is also moderate expression in kidney tissues and lung cell lines.

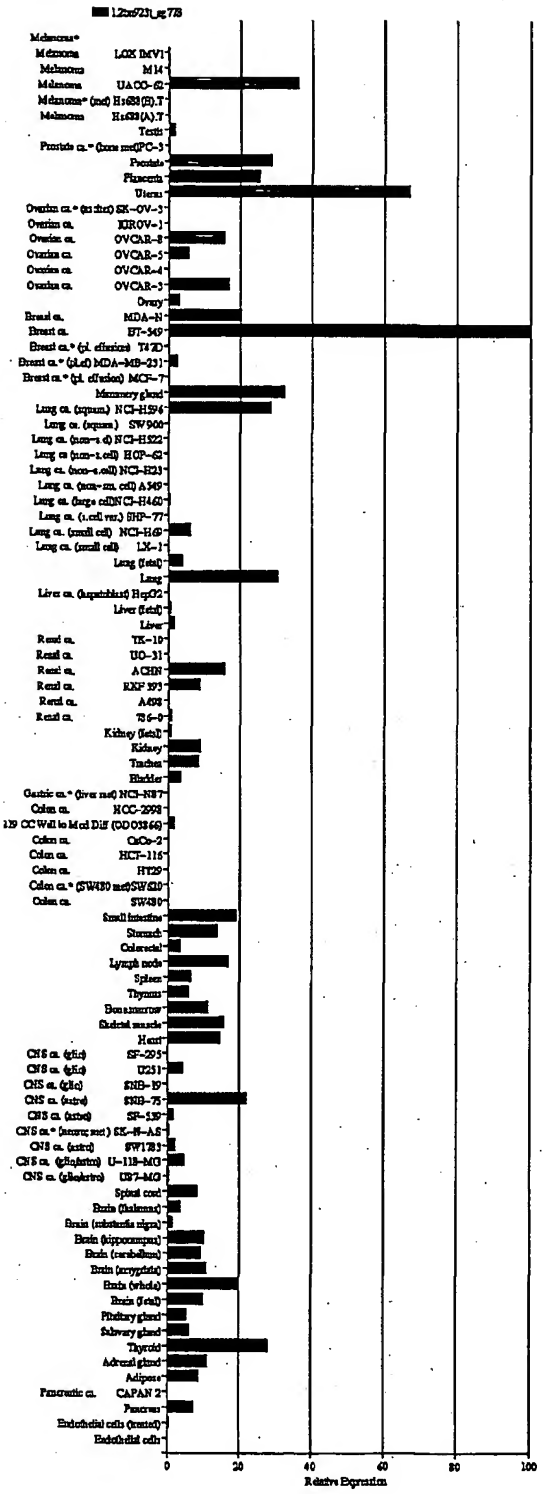
Panel 2.2 Summary: Ag1215 There appears to be widespread expression of the MOL5a gene in the samples of panel 2.2. Specifically, there seems to be an association of expression in breast cancer and normal ovarian tissue. This is reasonably consistent with the results obtained from Panel 1.2. In addition, there is also some correlation with expression in normal kidney tissue when compared to kidney cancers, also consistent with the observations in Panel 1.2. Thus, therapeutic modulation of this gene or gene product might show utility in the treatment of breast cancer, ovarian cancer or kidney cancer.

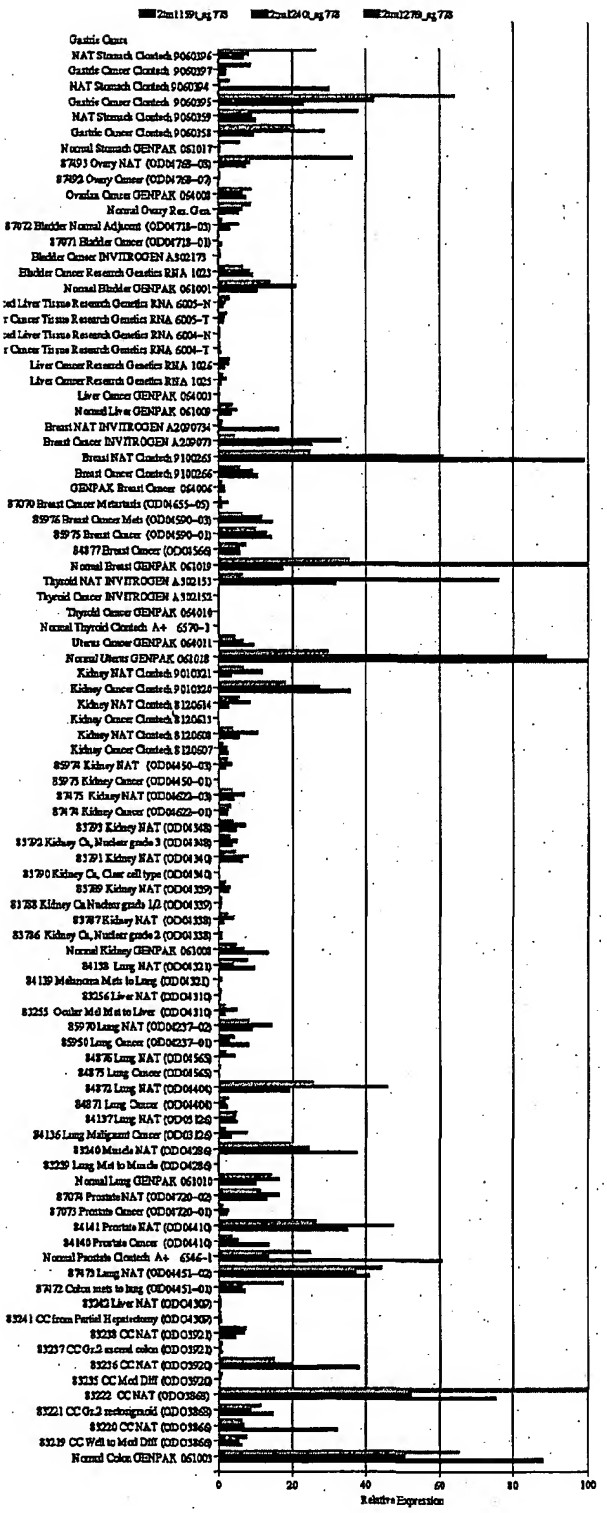
Panel 4D Summary: Ag1215/Ag1382 Results from two replicate experiments performed using probe and primer sets of identical sequences are in reasonable agreement. The MOL5a transcript is widely expressed in cell lines from this panel (CT values = 25-30), including thymus, lung, muco-epidermoid cell lines, fibroblasts from diverse origin, and activated T cells. In addition, the MOL5a gene is expressed in normal colon but not in colons from patients with Crohn's disease or colitis. Thus, protein therapeutics designed with the putative semaphorin encoded for by this protein could reduce or eliminate inflammation and tissue destruction due to IBD. High expression of this transcript was found on primary resting Th1 T cells, and also primary resting Th2 and Tr1 T cells. The high expression of this transcript in secondary T cells treated with CD95 suggests that this transcript encodes for a protein involved in activation of cell death. Furthermore, high expression of the MOL5A transcript is also found in activated basophils and eosinophils, suggesting a role for this protein in allergic disorder such as asthma, contact hypersensitivity, and hypersensitive immediate reactions. Antibody or protein therapeutics designed against the protein encoded for by this transcript could therefore reduce or inhibit inflammation in allergy, asthma, emphysema, psoriasis and/or autoimmunity.

Panel CNSD.01 Summary: Ag1215 Semaphorins can act as axon guidance proteins, specifically through their ability to act as chemorepellents that inhibit CNS regenerative capacity. Manipulation of levels of the MOL4 semaphorin-like protein may therefore be of use in inducing a compensatory synaptogenic response to neuronal death in Alzheimer's disease, Parkinson's disease, Huntington's disease, spinocerebellar ataxia, progressive supranuclear palsy, multiple sclerosis, ALS, head trauma, stroke, or any other disease/condition associated with neuronal loss.

Example 2: TaqMan Data for MOL7

TaqMan data was acquired for MOL7 as described in Eaxample 1 using the primers specified. The relative expression of MOL7 in the described tissues is represented in the graphs below.


LOX IMVI
M14
UACC-62
Testis
Prostate
Placenta
Uterus
OVCAR-8
OVCAR-5
OVCAR-4
OVCAR-3
MDA-N
BT-549
MCF-7
Mammary gland
NCI-H596
SW900
NCI-H522
HOP-62
NCI-H23
A549
NCI-H460
SHP-77
NCI-H69
LX-1
Lung (fetal)
Lung
HepG2
Liver (fetal)
Liver
TK-10
UO-31
ACHN
RXF 393
A498
786-0
Kidney (fetal)
Kidney
Trachea
Bladder
NCI-N87
HCC-2998
CaCo-2
HCT-116
HT29
SW620
SW480
Small intestine
Stomach
Colorectal
Lymph node
Spleen
Thymus
Bone marrow
Skeletal muscle
Heart
SF-295
U251
SNB-19
SNB-75
SF-539
SK-N-AS
SW1783
U-118-MG
U87-MG
Spinal cord
Brain (thalamus)
Brain (substantia nigra)
Brain (hippocampus)
Brain (cerebellum)
Brain (amygdala)
Brain (whole)
Brain (fetal)
Pituitary gland
Salivary gland
Thyroid
Adrenal gland
Adipose
CAPAN 2
Pancreas
Endothelial cells (renal)
Endothelial cells
0
20
40
60
80
100
Relative Expression


2xn1159_xg78
2xn1240_xg78
2xn1279_xg78
Gastric Cancer
NAT Stomach Clontech 9060396
Gastric Cancer Clontech 9060397
NAT Stomach Clontech 9060394
Gastric Cancer Clontech 9060395
NAT Stomach Clontech 9060359
Gastric Cancer Clontech 9060358
Normal Stomach GENPAK 061017
87493 Ovary NAT (ODD04768-08)
87492 Ovary Cancer (ODD04768-07)
Ovarian Cancer GENPAK 064008
Normal Ovary Res. Gen
87072 Bladder Normal Adjacent (ODD04718-03)
87071 Bladder Cancer (ODD04718-01)
Bladder Cancer INVITROGEN A302173
Bladder Cancer Research Genetics RNA 1023
Normal Bladder GENPAK 061001
Liver Tissue Research Genetics RNA 6005-N
Cancer Tissue Research Genetics RNA 6005-T
Liver Tissue Research Genetics RNA 6004-N
Cancer Tissue Research Genetics RNA 6004-T
Liver Cancer Research Genetics RNA 1026
Liver Cancer Research Genetics RNA 1025
Liver Cancer GENPAK 064003
Normal Liver GENPAK 061009
Breast NAT INVITROGEN A209734
Breast Cancer INVITROGEN A209073
Breast NAT Clontech 9100265
Breast Cancer Clontech 9100266
GENPAK Breast Cancer 064006
87070 Breast Cancer Metastatic (ODD04655-05)
85976 Breast Cancer Mets (ODD04590-03)
85975 Breast Cancer (ODD04590-01)
84877 Breast Cancer (ODD03568)
Normal Breast GENPAK 061019
Thyroid NAT INVITROGEN A302153
Thyroid Cancer INVITROGEN A302152
Thyroid Cancer GENPAK 064010
Normal Thyroid Clontech A+ 6570-1
Uterus Cancer GENPAK 064011
Normal Uterus GENPAK 061018
Kidney NAT Clontech 9010321
Kidney Cancer Clontech 9010320
Kidney NAT Clontech 8120614
Kidney Cancer Clontech 8120613
Kidney NAT Clontech 8120608
Kidney Cancer Clontech 8120607
85974 Kidney NAT (ODD04450-03)
85973 Kidney Cancer (ODD04450-01)
87475 Kidney NAT (ODD04622-03)
87474 Kidney Cancer (ODD04622-01)
83793 Kidney NAT (ODD04338)
83792 Kidney Ca, Nuclear grade 3 (ODD04338)
83791 Kidney NAT (ODD04340)
83790 Kidney Ca, Clear cell type (ODD04340)
83789 Kidney NAT (ODD04339)
83788 Kidney Ca Nuclear grade 1/2 (ODD04339)
83787 Kidney NAT (ODD04338)
83786 Kidney Ca, Nuclear grade 2 (ODD04338)
Normal Kidney GENPAK 061008
84138 Lung NAT (ODD04321)
84139 Melanoma Met to Lung (ODD04321)
83256 Liver NAT (ODD04310)
83255 Ocular Mel Met to Liver (ODD04310)
85970 Lung NAT (ODD04237-02)
85950 Lung Cancer (ODD04237-01)
84876 Lung NAT (ODD04565)
84875 Lung Cancer (ODD04565)
84872 Lung NAT (ODD04404)
84871 Lung Cancer (ODD04404)
84137 Lung NAT (ODD03126)
84136 Lung Malignant Cancer (ODD03126)
83240 Muscle NAT (ODD04286)
83239 Lung Met to Muscle (ODD04286)
Normal Lung GENPAK 061010
87078 Prostate NAT (ODD04720-02)
87073 Prostate Cancer (ODD04720-01)
84141 Prostate NAT (ODD04410)
84140 Prostate Cancer (ODD04410)
Normal Prostate Clontech A+ 6546-1
87473 Lung NAT (ODD04451-02)
87472 Colon mets to lung (ODD04451-01)
83242 Liver NAT (ODD04309)
83241 CC from Partial Hepatectomy (ODD04309)
83238 CC NAT (ODD03921)
83237 CC Gr.2 ascend colon (ODD03921)
83236 CC NAT (ODD03920)
83235 CC Mod Diff (ODD03920)
83222 CC NAT (ODD03868)
83221 CC Gr.2 rectosigmoid (ODD03868)
83220 CC NAT (ODD03869)
83219 CC Well to Mod Diff (ODD03869)
Normal Colon GENPAK 061003
0
20
40
60
80
100
Relative Expression

Example 3 SeqCallingTM Technology

cDNA was derived from various human samples representing multiple tissue types, normal and diseased states, physiological states, and developmental states from different donors. Samples were obtained as whole tissue, cell lines, primary cells or tissue cultured primary cells and cell lines. Cells and cell lines may have been treated with biological or chemical agents that regulate gene expression for example, growth factors, chemokines, steroids. The cDNA thus derived was then sequenced using CuraGen's proprietary SeqCalling technology. Sequence traces were evaluated manually and edited for corrections if appropriate. cDNA sequences from all samples were assembled with themselves and with public ESTs using bioinformatics programs to generate CuraGen's human SeqCalling database of SeqCalling assemblies. Each assembly contains one or more overlapping cDNA sequences derived from one or more human samples. Fragments and ESTs were included as components for an assembly when the extent of identity with another component of the assembly was at least 95% over 50 bp. Each assembly can represent a gene and/or its variants such as splice forms and/or single nucleotide polymorphisms (SNPs) and their combinations.

Variant sequences are included in this application. A variant sequence can include a single nucleotide polymorphism (SNP). A SNP can, in some instances, be referred to as a "cSNP" to denote that the nucleotide sequence containing the SNP originates as a cDNA. A SNP can arise in several ways. For example, a SNP may be due to a substitution of one nucleotide for another at the polymorphic site. Such a substitution can be either a transition or a transversion. A SNP can also arise from a deletion of a nucleotide or an insertion of a nucleotide, relative to a reference allele. In this case, the polymorphic site is a site at which one allele bears a gap with respect to a particular nucleotide in another allele. SNPs occurring within genes may result in an alteration of the amino acid encoded by the gene at the position of the SNP. Intragenic SNPs may also be silent, however, in the case that a codon including a SNP encodes the same amino acid as a result of the redundancy of the genetic code. SNPs occurring outside the region of a gene, or in an intron within a gene, do not result in changes in any amino acid sequence of a protein but may result in altered regulation of the expression pattern for example, alteration in temporal expression, physiological response regulation, cell type expression regulation, intensity of expression, stability of transcribed message.

Method of novel SNP Identification: SNPs are identified by analyzing sequence assemblies using CuraGen's proprietary SNPTool algorithm. SNPTool identifies variation in assemblies with the following criteria: SNPs are not analyzed within 10 base pairs on both

ends of an alignment; Window size (number of bases in a view) is 10; The allowed number of mismatches in a window is 2; Minimum SNP base quality (PHRED score) is 23; Minimum number of changes to score an SNP is 2/assembly position. SNPTool analyzes the assembly and displays SNP positions, associated individual variant sequences in the assembly, the depth of the assembly at that given position, the putative assembly allele frequency, and the SNP sequence variation. Sequence traces are then selected and brought into view for manual validation. The consensus assembly sequence is imported into CuraTools along with variant sequence changes to identify potential amino acid changes resulting from the SNP sequence variation. Comprehensive SNP data analysis is then exported into the SNPCalling database.

Method of novel SNP Confirmation: SNPs are confirmed employing a validated method know as Pyrosequencing (Pyrosequencing, Westborough, MA). Detailed protocols for Pyrosequencing can be found in:

Alderborn et al. Determination of Single Nucleotide Polymorphisms by Real-time Pyrophosphate DNA Sequencing. (2000). Genome Research. 10, Issue 8, August. 1249-1265.

In brief, Pyrosequencing is a real time primer extension process of genotyping. This protocol takes double-stranded, biotinylated PCR products from genomic DNA samples and binds them to streptavidin beads. These beads are then denatured producing single stranded bound DNA. SNPs are characterized utilizing a technique based on an indirect bioluminometric assay of pyrophosphate (PPi) that is released from each dNTP upon DNA chain elongation. Following Klenow polymerase-mediated base incorporation, PPi is released and used as a substrate, together with adenosine 5'-phosphosulfate (APS), for ATP sulfurylase, which results in the formation of ATP. Subsequently, the ATP accomplishes the conversion of luciferin to its oxi-derivative by the action of luciferase. The ensuing light output becomes proportional to the number of added bases, up to about four bases. To allow processivity of the method dNTP excess is degraded by apyrase, which is also present in the starting reaction mixture, so that only dNTPs are added to the template during the sequencing. The process has been fully automated and adapted to a 96-well format, which allows rapid screening of large SNP panels.

EQUIVALENTS

Although particular embodiments have been disclosed herein in detail, this has been done by way of example for purposes of illustration only, and is not intended to be limiting with respect to the scope of the appended claims, which follow. In particular, it is contemplated by the inventors that various substitutions, alterations, and modifications may be made to the invention without departing from the spirit and scope of the invention as defined by the claims. The choice of nucleic acid starting material, clone of interest, or library type is believed to be a matter of routine for a person of ordinary skill in the art with knowledge of the embodiments described herein. Other aspects, advantages, and modifications considered to be within the scope of the following claims.

WHAT IS CLAIMED IS:

1. An isolated polypeptide comprising an amino acid sequence selected from the group consisting of:

(a) a mature form of an amino acid sequence selected from the group consisting of SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, and 24;

(b) a variant of a mature form of an amino acid sequence selected from the group consisting of SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, and 24, wherein one or more amino acid residues in said variant differs from the amino acid sequence of said mature form, provided that said variant differs in no more than 15% of the amino acid residues from the amino acid sequence of said mature form;

(c) an amino acid sequence selected from the group consisting of SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, and 24; and

(d) a variant of an amino acid sequence selected from the group consisting of SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, and 24, wherein one or more amino acid residues in said variant differs from the amino acid sequence of said mature form, provided that said variant differs in no more than 15% of amino acid residues from said amino acid sequence.

2 The polypeptide of claim 1, wherein said polypeptide comprises the amino acid sequence of a naturally-occurring allelic variant of an amino acid sequence selected from the group consisting of SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, and 24.

3. The polypeptide of claim 2, wherein said allelic variant comprises an amino acid sequence that is the translation of a nucleic acid sequence differing by a single nucleotide from a nucleic acid sequence selected from the group consisting of SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, and 23.

4. The polypeptide of claim 1, wherein the amino acid sequence of said variant comprises a conservative amino acid substitution.

5. An isolated nucleic acid molecule comprising a nucleic acid sequence encoding a polypeptide comprising an amino acid sequence selected from the group consisting of:

(a) a mature form of an amino acid sequence selected from the group consisting of SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, and 24;

(b) a variant of a mature form of an amino acid sequence selected from the group consisting of SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, and 24, wherein one or more amino acid residues in said variant differs from the amino acid sequence of said mature form, provided that said variant differs in no more than 15% of the amino acid residues from the amino acid sequence of said mature form;

(c) an amino acid sequence selected from the group consisting of SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, and 24;

(d) a variant of an amino acid sequence selected from the group consisting SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, and 24, wherein one or more amino acid residues in said variant differs from the amino acid sequence of said mature form, provided that said variant differs in no more than 15% of amino acid residues from said amino acid sequence;

(e) a nucleic acid fragment encoding at least a portion of a polypeptide comprising an amino acid sequence chosen from the group consisting of SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, and 24, or a variant of said polypeptide, wherein one or more amino acid residues in said variant differs from the amino acid sequence of said mature form, provided that said variant differs in no more than 15% of amino acid residues from said amino acid sequence; and

(f) a nucleic acid molecule comprising the complement of (a), (b), (c), (d) or (e).

6. The nucleic acid molecule of claim 5, wherein the nucleic acid molecule comprises the nucleotide sequence of a naturally-occurring allelic nucleic acid variant.

7. The nucleic acid molecule of claim 5, wherein the nucleic acid molecule encodes a polypeptide comprising the amino acid sequence of a naturally-occurring polypeptide variant.

8. The nucleic acid molecule of claim 5, wherein the nucleic acid molecule differs by a single nucleotide from a nucleic acid sequence selected from the group consisting of SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, and 23.

9. The nucleic acid molecule of claim 5, wherein said nucleic acid molecule comprises a nucleotide sequence selected from the group consisting of:
   (a) a nucleotide sequence selected from the group consisting of SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, and 23;
   (b) a nucleotide sequence differing by one or more nucleotides from a nucleotide sequence selected from the group consisting of SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, and 23, provided that no more than 20% of the nucleotides differ from said nucleotide sequence;
   (c) a nucleic acid fragment of (a); and
   (d) a nucleic acid fragment of (b).

10. The nucleic acid molecule of claim 5, wherein said nucleic acid molecule hybridizes under stringent conditions to a nucleotide sequence chosen from the group consisting of SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, and 23, or a complement of said nucleotide sequence.

11. The nucleic acid molecule of claim 5, wherein the nucleic acid molecule comprises a nucleotide sequence selected from the group consisting of:
   (a) a first nucleotide sequence comprising a coding sequence differing by one or more nucleotide sequences from a coding sequence encoding said amino acid sequence, provided that no more than 20% of the nucleotides in the coding sequence in said first nucleotide sequence differ from said coding sequence;
   (b) an isolated second polynucleotide that is a complement of the first polynucleotide; and
   (c) a nucleic acid fragment of (a) or (b).

12. A vector comprising the nucleic acid molecule of claim 11.

13. The vector of claim 12, further comprising a promoter operably-linked to said nucleic acid molecule.

14. A cell comprising the vector of claim 12.

15. An antibody that binds immunospecifically to the polypeptide of claim 1.

16. The antibody of claim 15, wherein said antibody is a monoclonal antibody.

17. The antibody of claim 15, wherein the antibody is a humanized antibody.

18. A method for determining the presence or amount of the polypeptide of claim 1 in a sample, the method comprising:

(a) providing the sample;

(b) contacting the sample with an antibody that binds immunospecifically to the polypeptide; and

(c) determining the presence or amount of antibody bound to said polypeptide,

thereby determining the presence or amount of polypeptide in said sample.

19. A method for determining the presence or amount of the nucleic acid molecule of claim 5 in a sample, the method comprising:

(a) providing the sample;

(b) contacting the sample with a probe that binds to said nucleic acid molecule; and

(c) determining the presence or amount of the probe bound to said nucleic acid molecule,

thereby determining the presence or amount of the nucleic acid molecule in said sample.

20. The method of claim 19 wherein presence or amount of the nucleic acid molecule is used as a marker for cell or tissue type.

21. The method of claim 20 wherein the cell or tissue type is cancerous.

22. A method of identifying an agent that binds to a polypeptide of claim 1, the method comprising:

(a) contacting said polypeptide with said agent; and

(b) determining whether said agent binds to said polypeptide.

23. The method of claim 22 wherein the agent is a cellular receptor or a downstream effector.

24. A method for identifying an agent that modulates the expression or activity of the polypeptide of claim 1, the method comprising:
(a) providing a cell expressing said polypeptide;
(b) contacting the cell with said agent, and
(c) determining whether the agent modulates expression or activity of said polypeptide,
whereby an alteration in expression or activity of said peptide indicates said agent modulates expression or activity of said polypeptide.

25. A method for modulating the activity of the polypeptide of claim 1, the method comprising contacting a cell sample expressing the polypeptide of said claim with a compound that binds to said polypeptide in an amount sufficient to modulate the activity of the polypeptide.

26. A method of treating or preventing a MOLX-associated disorder, said method comprising administering to a subject in which such treatment or prevention is desired the polypeptide of claim 1 in an amount sufficient to treat or prevent said MOLX-associated disorder in said subject.

27. The method of claim 26 wherein the disorder is selected from the group consisting of cardiomyopathy and atherosclerosis.

28. The method of claim 26 wherein the disorder is related to cell signal processing and metabolic pathway modulation.

29. The method of claim 26, wherein said subject is a human.

30. A method of treating or preventing a MOLX-associated disorder, said method comprising administering to a subject in which such treatment or prevention is desired

the nucleic acid of claim 5 in an amount sufficient to treat or prevent said MOLX-associated disorder in said subject.

31. The method of claim 30 wherein the disorder is selected from the group consisting of cardiomyopathy and atherosclerosis.

32. The method of claim 30 wherein the disorder is related to cell signal processing and metabolic pathway modulation.

33. The method of claim 30, wherein said subject is a human.

34. A method of treating or preventing a MOLX-associated disorder, said method comprising administering to a subject in which such treatment or prevention is desired the antibody of claim 15 in an amount sufficient to treat or prevent said MOLX-associated disorder in said subject.

35. The method of claim 34 wherein the disorder is diabetes.

36. The method of claim 34 wherein the disorder is related to cell signal processing and metabolic pathway modulation.

37. The method of claim 34, wherein the subject is a human.

38. A pharmaceutical composition comprising the polypeptide of claim 1 and a pharmaceutically-acceptable carrier.

39. A pharmaceutical composition comprising the nucleic acid molecule of claim 5 and a pharmaceutically-acceptable carrier.

40. A pharmaceutical composition comprising the antibody of claim 15 and a pharmaceutically-acceptable carrier.

41. A kit comprising in one or more containers, the pharmaceutical composition of claim 38.

42. A kit comprising in one or more containers, the pharmaceutical composition of claim

43. A kit comprising in one or more containers, the pharmaceutical composition of claim 4

44. A method for determining the presence of or predisposition to a disease associated with altered levels of the polypeptide of claim 1 in a first mammalian subject, the method comprising:

(a) measuring the level of expression of the polypeptide in a sample from the first mammalian subject; and

(b) comparing the amount of said polypeptide in the sample of step (a) to the amount of the polypeptide present in a control sample from a second mammalian subject known not to have, or not to be predisposed to, said disease;

wherein an alteration in the expression level of the polypeptide in the first subject as compared to the control sample indicates the presence of or predisposition to said disease.

45. The method of claim 44 wherein the predisposition is to cancers.

46. A method for determining the presence of or predisposition to a disease associated with altered levels of the nucleic acid molecule of claim 5 in a first mammalian subject, the method comprising:

(a) measuring the amount of the nucleic acid in a sample from the first mammalian subject; and

(b) comparing the amount of said nucleic acid in the sample of step (a) to the amount of the nucleic acid present in a control sample from a second mammalian subject known not to have or not be predisposed to, the disease;

wherein an alteration in the level of the nucleic acid in the first subject as compared to the control sample indicates the presence of or predisposition to the disease.

47. The method of claim 46 wherein the predisposition is to a cancer.

48. A method of treating a pathological state in a mammal, the method comprising administering to the mammal a polypeptide in an amount that is sufficient to alleviate the pathological state, wherein the polypeptide is a polypeptide having an amino acid sequence at least 95% identical to a polypeptide comprising an amino acid sequence of at least one of SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, and 24, or a biologically active fragment thereof.

49. A method of treating a pathological state in a mammal, the method comprising administering to the mammal the antibody of claim 15 in an amount sufficient to alleviate the pathological state.

50. A method for the screening of a candidate substance interacting with an olfactory receptor polypeptide selected from the group consisting of SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, and 24, or fragments or variants thereof, comprises the following steps:

a) providing a polypeptide selected from the group consisting of the sequences of SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, and 24, or a peptide fragment or a variant thereof;

b) obtaining a candidate substance;

c) bringing into contact said polypeptide with said candidate substance; and

d) detecting the complexes formed between said polypeptide and said candidate substance.

51. A method for the screening of ligand molecules interacting with an olfactory receptor polypeptide selected from the group consisting of SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, and 24, wherein said method comprises:

a) providing a recombinant eukaryotic host cell containing a nucleic acid encoding a polypeptide selected from the group consisting of the polypeptides comprising the amino acid sequences SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, and 24;

b) preparing membrane extracts of said recombinant eukaryotic host cell;

c) bringing into contact the membrane extracts prepared at step b) with a selected ligand molecule; and

d) detecting the production level of second messengers metabolites.

52. A method for the screening of ligand molecules interacting with an olfactory receptor polypeptide selected from the group consisting of SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, and 24, wherein said method comprises:

a) providing an adenovirus containing a nucleic acid encoding a polypeptide selected from the group consisting of polypeptides comprising the amino acid sequences SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, and 24;

b) infecting an olfactory epithelium with said adenovirus;

c) bringing into contact the olfactory epithelium b) with a selected ligand molecule; and

d) detecting the increase of the response to said ligand molecule.

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
24 January 2002 (24.01.2002)

PCT

(10) International Publication Number
WO 02/006339 A3

(51) International Patent Classification[7]: C12N 15/00, C07K 14/47, C12Q 1/68, G01N 33/53, 33/50, C07K 16/18, A61K 38/17

(21) International Application Number: PCT/US01/21249

(22) International Filing Date: 3 July 2001 (03.07.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/215,854	3 July 2000 (03.07.2000)	US
60/215,856	3 July 2000 (03.07.2000)	US
60/215,902	3 July 2000 (03.07.2000)	US
60/216,585	7 July 2000 (07.07.2000)	US
60/216,586	7 July 2000 (07.07.2000)	US
60/216,722	7 July 2000 (07.07.2000)	US
60/218,622	17 July 2000 (17.07.2000)	US
60/218,992	17 July 2000 (17.07.2000)	US
60/221,285	27 July 2000 (27.07.2000)	US
60/268,734	14 February 2001 (14.02.2001)	US
60/274,260	8 March 2001 (08.03.2001)	US
60/279,856	29 March 2001 (29.03.2001)	US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier applications:

US	60/215,854 (CIP)
Filed on	3 July 2000 (03.07.2000)
US	60/215,902 (CIP)
Filed on	3 July 2000 (03.07.2000)
US	60/215,856 (CIP)
Filed on	3 July 2000 (03.07.2000)
US	60/216,585 (CIP)
Filed on	7 July 2000 (07.07.2000)
US	60/216,586 (CIP)
Filed on	7 July 2000 (07.07.2000)
US	60/216,722 (CIP)
Filed on	7 July 2000 (07.07.2000)
US	60/218,622 (CIP)
Filed on	17 July 2000 (17.07.2000)
US	60/218,992 (CIP)
Filed on	17 July 2000 (17.07.2000)
US	60/221,285 (CIP)
Filed on	27 July 2000 (27.07.2000)
US	60/268,734 (CIP)
Filed on	14 February 2001 (14.02.2001)
US	60/274,260 (CIP)
Filed on	8 March 2001 (08.03.2001)
US	60/279,856 (CIP)
Filed on	29 March 2001 (29.03.2001)

(71) Applicant *(for all designated States except US)*: **CURAGEN CORPORATION** [US/US]; 555 Long Wharf Drive, 11th Floor, New Haven, CT 06511 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **SPADERNA, Steven, K.** [US/US]; 261 Deerfield Drive, Berlin, CT 06037 (US). **TCHERNEV, Velizar** [BG/US]; 45 Jefferson Road #3-12, Branford, CT 06405 (US). **LIU, Xiaohong** [US/US]; 360 Neponset Street, Apartment 509, Canton, MA 02021 (US). **SHENOY, Suresh** [IN/US]; 15 Millwood Drive, Branford, CT 06405 (US). **SPYTEK, Kimberly** [US/US]; 28 Court Street #1, New Haven, CT 06511 (US). **ZERHUSEN, Bryan** [US/US]; 337 Monticello Drive, Branford, CT 06405 (US). **PATTURAJAN, Meera** [IN/US]; 45 Harrison Avenue, Apartment 1C, Branford, CT 06405 (US). **TAUPIER, Raymond, J.** [US/US]; 34 Pardee Place Extension, East Haven, CT 06512 (US). **RASTELLI, Luca** [IT/US]; 52 Pepperbush Lane, Guilford, CT 06437 (US). **GROSSE, William, M.** [US/US]; 15 Rice Terrace Road, Branford, CT 06405 (US). **SZEKERES, Edward, S.** [US/US]; 110 Constitution Street, Wallingford, CT 06492 (US). **ALSOBROOK, John, II** [US/US]; 60 Lake Drive, Madison, CT 06443 (US). **LEPLEY, Denise, M.** [US/US]; 51 Church Street, Branford, CT 06405 (US). **SHEN, Lei** [US/US]; 109 Kaye Vue Drive, Apartment 3K, Hamden, CT 06514 (US). **BURGESS, Catherine, E.** [US/US]; 90 Carriage Hill Drive, Wethersfield, CT 06109 (US). **SHIMKETS, Richard, A.** [US/US]; 5 Indian Meadows Drive, Guilford, Connecticut 06437 (US). **PADIGARU, Muralidhara** [IN/US]; 71 Hampton Park, Branford, CT 06405 (US).

(74) Agent: **ELRIFI, Ivor, R.**; Mintz, Levin, Cohn, Ferris, Glovsky and Popeo, P.C., One Financial Center, Boston, MA 02111 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX,

[Continued on next page]

(54) Title: PROTEINS AND NUCLEIC ACIDS ENCODING SAME

(57) Abstract: Disclosed herein are nucleic acid sequences that encode novel polypeptides. Also disclosed are polypeptides encoded by these nucleic acid sequences, and antibodies, which immunospecifically-bind to the polpeptide, as well as derivatives, variants, mutants, or fragments of the aforementioned polypeptide, polynucleotide, or antibody. The invention further discloses therapeutic, diagnostic and research methods for diagnosis, treatment, and prevention of disorders involving any one of these novel human nucleic acids and proteins.

WO 02/006339 A3

MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) **Designated States** *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*

(88) **Date of publication of the international search report:**
31 July 2003

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

INTERNATIONAL SEARCH REPORT

National Application No
PCT/US 01/21249

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 C12N15/00 C07K14/47 C12Q1/68 G01N33/53 G01N33/50
C07K16/18 A61K38/17

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 C12N C07K C12Q G01N A61K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)
EPO-Internal, MEDLINE, WPI Data, PAJ, BIOSIS, SEQUENCE SEARCH

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 94 07474 A (ARTAVANIS TSAKONAS SPYRIDON ;BLAUMUELLER CHRISTINE MARIE (US); FEH) 14 April 1994 (1994-04-14) 98.9%;99.9% identity to Seq. Id. Nos. 1/2;3/4 respectively figure 17 the whole document	1-52
X	WO 95 19779 A (UNIV YALE) 27 July 1995 (1995-07-27) 98.9%;99.9% identity to Seq. Id. Nos. 1/2;3/4 respectively figure 17 the whole document	1-52

-/--

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier document but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
24 September 2002	24.01.03

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL – 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Petri, B

Form PCT/ISA/210 (second sheet) (July 1992)

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication,where appropriate, of the relevant passages	Relevant to claim No.
A	STIFANI S ET AL: "HUMAN HOMOLOGS OF A DROSOPHILA ENHANCER OF SPLIT GENE PRODUCT DEFINE A NOVEL FAMILY OF NUCLEAR PROTEINS" NATURE GENETICS, NEW YORK, NY, US, vol. 2, no. 2, 1 October 1992 (1992-10-01), pages 119-127, XP000618720 ISSN: 1061-4036 ---	
A	ENCINAS J A ET AL: "Cloning, expression, and genetic mapping of Sema W, a member of the semaphorin family." PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA. UNITED STATES 2 MAR 1999, vol. 96, no. 5, 2 March 1999 (1999-03-02), pages 2491-2496, XP002214405 ISSN: 0027-8424 ---	
A	SEKIDO Y ET AL: "HUMAN SEMAPHORINS A(V) AND IV RESIDE IN THE 3P21.3 SMALL CELL LUNG CANCER DELETION REGION AND DEMONSTRATE DISTINCT EXPRESSION PATTERNS" PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF USA, NATIONAL ACADEMY OF SCIENCE. WASHINGTON, US, vol. 93, no. 9, April 1996 (1996-04), pages 4120-4125, XP000996622 ISSN: 0027-8424 ---	
A	ZLOBIN A ET AL: "TOWARD THE RATIONAL DESIGN OF CELL FATE MODIFIERS: NOTCH SIGNALING AS A TARGET FOR NOVEL BIOPHARMACEUTICALS" CURRENT PHARMACEUTICAL BIOTECHNOLOGY, BENTHAM SCIENCE PUBLISHERS, BOCA RATON,FL, US, vol. 1, no. 1, July 2000 (2000-07), pages 83-106, XP008004456 ISSN: 1389-2010 -----	

INTERNATIONAL SEARCH REPORT

nternational application No.
PCT/US 01/21249

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. [X] Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

Although claims 26-37, 48-49 are directed to a method of treatment of the human/animal body, the search has been carried out and based on the alleged effects of the compound/composition.

2. [X] Claims Nos.: 25 (partially)
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:

see FURTHER INFORMATION sheet PCT/ISA/210

3. [ ] Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

see additional sheet

1. [ ] As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.

2. [ ] As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. [ ] As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:

4. [X] No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

1-24, 26-52 (all partially)

Remark on Protest

[ ] The additional search fees were accompanied by the applicant's protest.

[ ] No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

Continuation of Box I.2

Claims Nos.: 25 (partially)

"Reach-through claim ", relates to the use of structurally undefined compounds in a method for modulating activity.

Present claim 25 relates to a method defined by reference to a desirable characteristic or property, namely "... a compound that binds to said polypeptide in an amount sufficient to modulate the activity of the polypeptide".

The claims cover all methods having this characteristic or property, whereas the application provides support within the meaning of Article 6 PCT and/or disclosure within the meaning of Article 5 PCT for none of such methods. In the present case, the claims so lack support, and the application so lacks disclosure, that a meaningful search over the whole of the claimed scope is impossible. Independent of the above reasoning, the claims also lack clarity (Article 6 PCT). An attempt is made to define the method by reference to a result to be achieved. Again, this lack of clarity in the present case is such as to render a meaningful search over the whole of the claimed scope impossible. Consequently, the search has been carried out for those parts of the claims which appear to be clear, supported and disclosed, namely those parts wherein the compound that binds to said polypeptide is an antibody (as disclosed on page 127 line 33 - page 128 line 1 of the specification).

The applicant's attention is drawn to the fact that claims, or parts of claims, relating to inventions in respect of which no international search report has been established need not be the subject of an international preliminary examination (Rule 66.1(e) PCT). The applicant is advised that the EPO policy when acting as an International Preliminary Examining Authority is normally not to carry out a preliminary examination on matter which has not been searched. This is the case irrespective of whether or not the claims are amended following receipt of the search report or during any Chapter II procedure.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. Claims: 1-24, 26-52 (all partially)

An isolated polypeptide comprising an amino acid sequence selected from the group consisting of: Seq. Id. No. 2, nucleic acids, vectors, hosts, antibodies, compositions, methods and uses relating thereto.

1.1. Claims: 1-24, 26-52 (all partially)
An isolated polypeptide comprising an amino acid sequence selected from the group consisting of: Seq. Id. No. 4 nucleic acids, vectors, hosts, antibodies, compositions, methods and uses relating thereto.

2. Claims: 1-24, 26-52 (all partially)

An isolated polypeptide comprising an amino acid sequence selected from the group consisting of: Seq. Id. No. 6, nucleic acids, vectors, hosts, antibodies, compositions, methods and uses relating thereto.

3. Claims: 1-24, 26-52 (all partially)

An isolated polypeptide comprising an amino acid sequence selected from the group consisting of: Seq. Id. No. 8/10, nucleic acids, vectors, hosts, antibodies, compositions, methods and uses relating thereto.

4. Claims: 1-24, 26-52 (all partially)

An isolated polypeptide comprising an amino acid sequence selected from the group consisting of: Seq. Id. No. 12, nucleic acids, vectors, hosts, antibodies, compositions, methods and uses relating thereto.

5. Claims: 1-24, 26-52 (all partially)

An isolated polypeptide comprising an amino acid sequence selected from the group consisting of: Seq. Id. No. 14, 16, 18 nucleic acids, vectors, hosts, antibodies, compositions, methods and uses relating thereto.

6. Claims: 1-24, 26-52 (all partially)

An isolated polypeptide comprising an amino acid sequence selected from the group consisting of: Seq. Id. No. 20, nucleic acids, vectors, hosts, antibodies, compositions,

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

methods and uses relating thereto.

7. Claims: 1-24, 26-52 (all partially)

An isolated polypeptide comprising an amino acid sequence selected from the group consisting of: Seq. Id. No. 22, nucleic acids, vectors, hosts, antibodies, compositions, methods and uses relating thereto.

8. Claims: 1-24, 26-52 (all partially)

An isolated polypeptide comprising an amino acid sequence selected from the group consisting of: Seq. Id. No. 24, nucleic acids, vectors, hosts, antibodies, compositions, methods and uses relating thereto.

Please note that all inventions mentioned under item 1, although not necessarily linked by a common inventive concept, could be searched without effort justifying an additional fee.

INTERNATIONAL SEARCH REPORT

Information on patent family members

tional Application No
PCT/US 01/21249

Patent document cited in search report	Publication date	Patent family member(s)		Publication date
WO 9407474 A	14-04-1994	US	5786158 A	28-07-1998
		AT	216225 T	15-05-2002
		AU	685067 B2	15-01-1998
		AU	5350394 A	26-04-1994
		CA	2145778 A1	14-04-1994
		DE	69331832 D1	23-05-2002
		DE	69331832 T2	19-12-2002
		DK	662827 T3	12-08-2002
		EP	1197220 A2	17-04-2002
		EP	1175909 A2	30-01-2002
		EP	0662827 A1	19-07-1995
		ES	2174855 T3	16-11-2002
		JP	8502170 T	12-03-1996
		PT	662827 T	30-09-2002
		WO	9407474 A1	14-04-1994
		US	6083904 A	04-07-2000
WO 9519779 A	27-07-1995	US	5750652 A	12-05-1998
		AU	1686395 A	08-08-1995
		WO	9519779 A1	27-07-1995